

Alma Mater Studiorum - Università di Bologna

SCUOLA DI SCIENZE

Corso di Laurea in

Informatica

Il caso Cambridge Analytica:

**Tecniche di profilazione e di microtargeting politico atte
alla segmentazione dell'elettorato americano**

Tesi di Laurea in

Strategia Aziendale

Relatore:

Prof. Edoardo Mollona

Laureanda:

Martina Bergonzoni

Matricola: 731162

Anno Accademico 2017/2018

Sessione I

INDICE

INTRODUZIONE	4
CAPITOLO 1: IL CASO CAMBRIDGE ANALYTICA	6
1.1 Attori in scena di questo celebre caso mediatico	6
1.1.1 Quali connessioni hanno i precedenti attori tra loro?	9
CAPITOLO 2: LA PROFILAZIONE	11
2.1 Tecniche di profilazione usate da CA	11
2.2 Teoria e fundamenta della profilazione attuata da Cambridge Analytica	15
2.2.1 Predizione di variabili numeriche e dicotomiche	19
2.3 Software per la profilazione: Apply Magic Sauce	23
2.3.1 La Prediction API di Apply Magic Sauce	26
2.4 Privacy policy: trattamento dei dati personali	32
CAPITOLO 3: MICROTARGETING POLITICO ATTUATO DA CA	34
3.1 Facebook come strumento d'influenza politica	34
3.2 Analisi scientifica di Kosinski: "Targeting psicologico come approccio efficace alla persuasione digitale di massa"	38
3.3 Tecnologie usate da Cambridge Analytica per influenzare il popolo americano targettizzato	43
CAPITOLO 4: SVILUPPI RECENTI	47
4.1 Innovazioni tecnologiche e cambiamenti in seguito allo scandalo di Cambridge Analytica	47
4.1.1 General Data Protection Regulation	51
CONCLUSIONI	58
BIBLIOGRAFIA	60
SITOGRAFIA	60
FIGURE	63
RINGRAZIAMENTI	64

INTRODUZIONE

In questa tesi si discuterà ampiamente del recente caso Cambridge Analytica-Facebook, in particolare della profilazione attuata ai danni di $\cong 50$ milioni di utenti Facebook, a fini elettorali, da parte della società britannica di consulenza politica che si occupava di data mining, web e social media analysis. Si affronterà il tema del meccanismo della manipolazione e persuasione dell'elettorato americano, che vede Facebook come uno strumento dal duplice utilizzo: profilazione e influenza successiva.

La Strategic Communication Laboratories, azienda madre di Cambridge Analytica, avvalendosi dell'applicazione Facebook "thisisyourdigitallife" ha estratto milioni di dati dagli utenti Facebook, per poi rivenderli all'azienda figlia, con i quali ha costruito accurati profili basati sulla psicomètria (indagine psicologica che attua una valutazione quantitativa dei comportamenti).

Una volta effettuata la profilazione delle diverse personalità, arrivare all'elettore, sfruttando punti deboli e insicurezze, risulta estremamente semplice. Il veicolo risulta essere sempre Facebook, non per l'estrazione dei dati, ma per influenzare quei profili creati su misura, attraverso le inserzioni customizzate e le fake news, allo scopo di danneggiare la concorrenza politica.

Nel primo capitolo si cercherà di fornire una visione generale degli attori coinvolti nello scandalo e le relazioni tra essi. Con il secondo capitolo, si entrerà nel vivo della questione, affrontando il tema della profilazione, basata su modelli predittivi (OCEAN: *Big Five*) della personalità e su uno studio effettuato da Michal Kosiski, psicologo e data scientist presso l'Università di Cambridge, il quale mostrò la possibilità di predire caratteristiche sensibili di un utente basandosi su un certo numero di "like" di Facebook. Si tratta di algoritmi di machine learning o apprendimento automatico da parte di software "intelligenti". La particolarità degli algoritmi di machine learning è che il software, la macchina, diventa tanto più intelligente ed efficiente quanto migliore è il set di dati a disposizione. È dimostrato che non è fondamentale la qualità del software, quanto piuttosto il training che gli algoritmi esercitano sui dati. Il fulcro principale e di estrema importanza risulta essere il dato. Maggiore è la quantità a disposizione e migliore sarà la capacità di anticipare il comportamento del cliente e fornire la soluzione alle sue necessità.

Il machine learning ha un approccio prettamente statistico e non affronta i temi semantici, filosofici o ontologici legati all'intelligenza artificiale. Si parlerà, in seguito, di Apply Magic Sauce, il software progettato e sviluppato da Kosinski e il suo team, che con un solo clic effettua una previsione del profilo personale, attraverso Facebook. Si analizzeranno le Prediction API del software in questione e se ne presenteranno esempi.

Per quanto riguarda il terzo capitolo, si affronterà il microtargeting politico effettuato da Cambridge Analytica, usufruendo di Facebook. Il *microtargeting* è una tipologia di marketing politico innovativo, fondato sulla misurazione della personalità degli elettori in base alle loro tracce digitali e nella pratica d'influenza attraverso l'invio di messaggi personalizzati. Verrà, inoltre, discusso uno studio dello stesso Kosinski, il quale dopo aver effettuato vari test in laboratorio ha finalmente risposto alla domanda: "il targeting psicologico è uno strumento efficace di propaganda digitale?". Infine, verranno introdotti gli strumenti di insight che ha sfruttato Cambridge Analytica, per segmentare e persuadere i suoi utenti bersaglio.

Nel quarto e ultimo capitolo, si elencheranno le novità, i cambiamenti e le innovazioni tecnologiche succedute allo scandalo. Verrà analizzato in maniera abbastanza approfondita il GDPR, General Data Protection Regulation, entrato in vigore il 25 maggio 2018, per cercare di regolamentare e limitare tale abuso di dati.

CAPITOLO 1

Il caso Cambridge Analytica

1.1 Attori in scena di questo celebre caso mediatico:

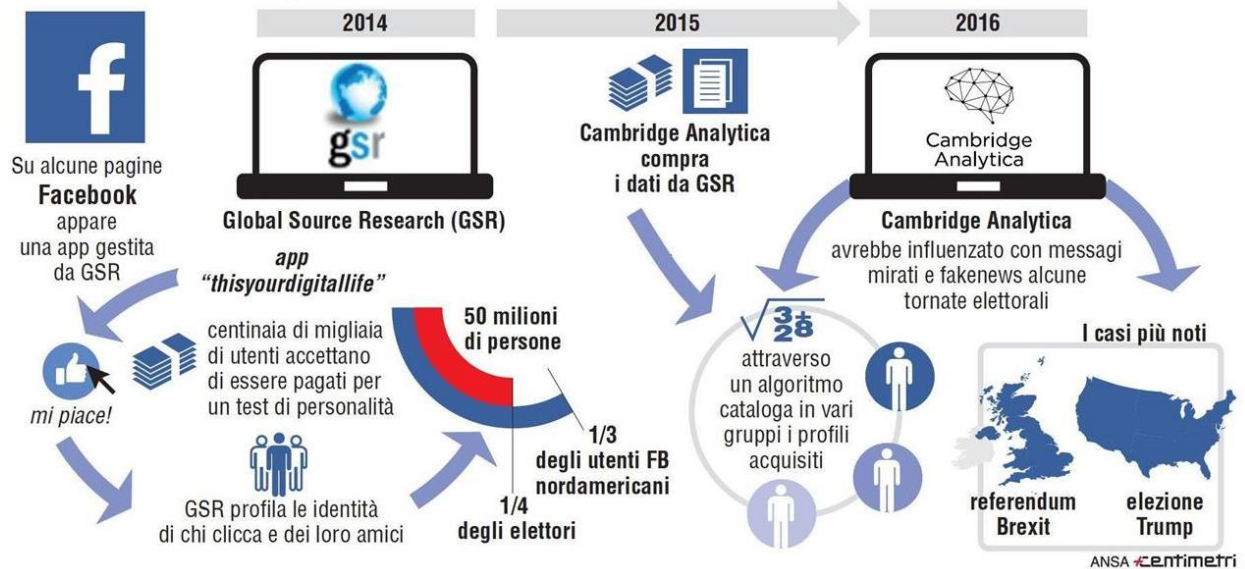


Figura 1.1: Visione rappresentativa del caso Cambridge Analytica,

<https://bit.ly/2u6ZitT>

- **Cambridge Analytica:**

società britannica di consulenza che fondeva l'analisi dei dati, il data mining e l'intermediazione dei dati con la comunicazione strategica per varie campagne elettorali;

- **Christopher Wylie:**

ex dipendente di Cambridge Analytica, ha svelato il meccanismo di raccolta dati da parte dell'azienda ai quotidiani Guardian e New York Times, che il 16 marzo 2018 hanno reso pubblico tale scandalo digitale;

- **Facebook:**

servizio web gratuito che permette alle persone di mettersi in contatto con amici vicini e lontani e di creare nuove relazioni. Essendo il social network più famoso del mondo, è in possesso dei dati di 2 miliardi di iscritti;

- **Global science research:**

società di proprietà dell'accademico russo-americano Aleksandr Kogan, che si occupa di ottimizzazione delle strategie di marketing attraverso il potere dei big data e delle scienze psicologiche;

- **Strategic Communication Laboratories:**

è una società privata di ricerca comportamentale britannica e società di comunicazione strategica. SCL ha formato Cambridge Analytica per partecipare al processo elettorale negli Stati Uniti;

- **Robert Mercer:**

fondatore della Cambridge Analytica nel 2013. Secondo Forbes, nel 2016, ha speso più di 25 milioni di dollari per sostenere la campagna di Trump attraverso un super PAC (political action committee) conservatore, Make America Number 1, presieduto dalla figlia Rebekah;

- **Steve Bannon:**

è la mente che ha portato alla nascita di Cambridge Analytica, inoltre è stato uno dei più fidati collaboratori di Donald Trump, coordinatore della campagna elettorale e, per parte del 2017, membro del Consiglio Nazionale per la Sicurezza.

- **Aleksandr Kogan:**

ricercatore a Cambridge, è un esperto in big data, analisi dei comportamenti sociali e neuroscienze. Fondatore della Global Science Research e sviluppatore dell'app quiz "thisisyourdigitallife". È l'uomo che ha materialmente raccolto e processato i dati personali di Facebook, cedendoli poi a CA;

- **Alexander Nix:**

è l'uomo che ha convinto Mercer e Bannon a creare l'azienda e a metterla a disposizione di alcune delle campagne politiche più importanti degli ultimi anni. Ha ricoperto il ruolo di CEO di Cambridge Analytica fino al 20 marzo, in quanto, travolto dallo scandalo, soprattutto per il video rilasciato da Channel 4, venne sospeso;

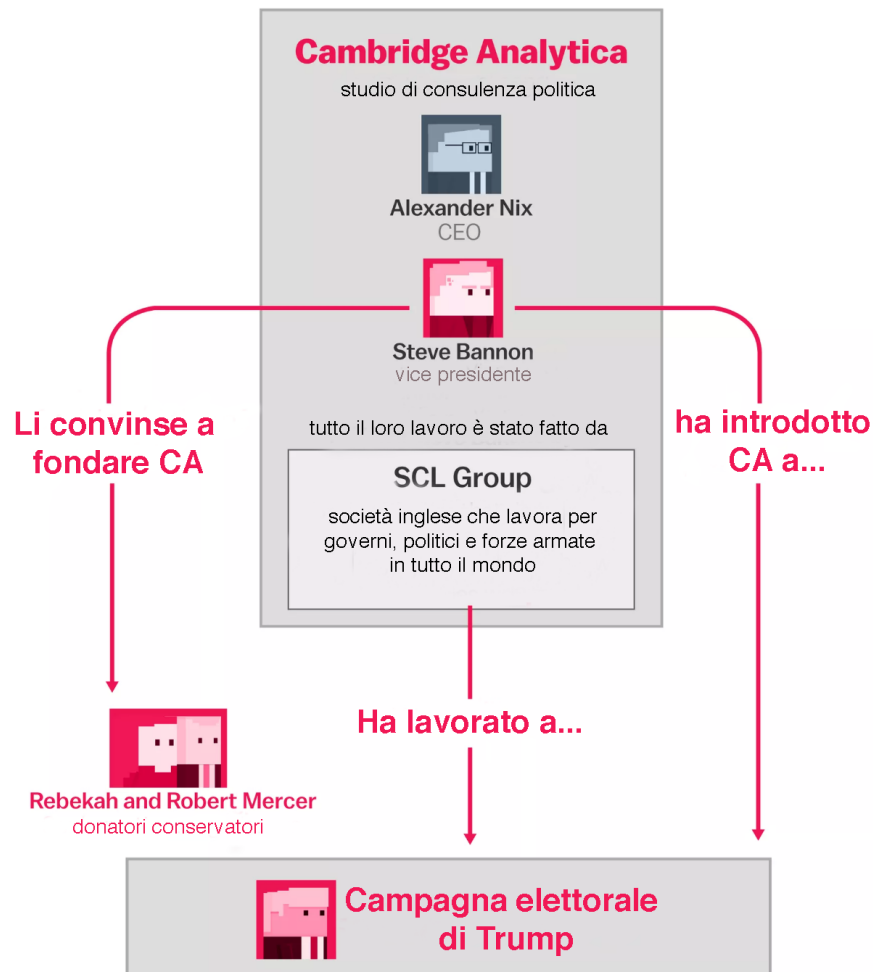


Figura 1.2: Organizzazione interna Cambridge Analytica, <https://bit.ly/2KSpQIK>

- **Michal Kosinski:**

è un ricercatore di Cambridge, psicologo e scienziato dei dati. Uno dei maggiori esperti di microtargeting comportamentale, ovvero di tutti quei sistemi che generano pubblicità personalizzata su ogni singolo utente, comprendendo anche la parte emotiva.

1.1.1 Quali connessioni hanno i precedenti attori tra loro?

Cambridge Analytica ha acquisito i dati di ≈ 50 milioni di utenti Facebook, solo negli Stati Uniti, tramite una terza società, la *Global science research* (Gsr). Quest'ultima, di proprietà dell'accademico russo-americano *Aleksandr Kogan*, nel 2014 avrebbe ottenuto informazioni pagando gli utenti per farsi sottoporre ad un quiz sulla personalità, che prometteva di produrre profili psicologici e di previsione del proprio comportamento, basandosi sulle attività svolte online. Per utilizzarla, gli utenti dovevano collegarsi utilizzando Facebook Login, il sistema che permette di iscriversi ad un sito senza la necessità di creare nuovi username e password, utilizzando invece la verifica controllata da Facebook. Il servizio era gratuito, ma come spesso avviene online, in realtà risultava "pagato" con i dati degli utenti: tale applicazione otteneva l'accesso all'indirizzo email, età, sesso e alle altre informazioni contenute nel proprio profilo. All'epoca il social network permetteva ai gestori delle applicazioni di raccogliere anche alcuni dati sulla rete di amici della persona appena iscritta. Difatti, Facebook ha consentito allo sviluppatore di progettare l'applicazione al solo scopo di raccogliere dati, ma esclusivamente a fini accademici. Kogan è stato in grado di individuare e sfruttare quest'ultima scappatoia, ottenendo accesso ai dati - seppure in forma aggregata ed anonima - di tutte le persone che hanno scaricato e usato tale app, ≈ 270 mila soggetti, ma anche di tutti i loro "amici", arrivando ad un totale di oltre 50 milioni di utenti.

Dopo lo scoppio dello scandalo, il titolo di Facebook è crollato a Wall Street per due giorni consecutivi, bruciando 9 miliardi di dollari. (Il 20 marzo ha registrato un calo del -5,18%). A pagare dazio è stato anche Twitter: lo scandalo, insomma, ha penalizzato l'intero settore dei social media, che vede arrivare all'orizzonte nuove norme e regolamentazioni.

Negli Stati Uniti, Cambridge Analytica è stata accusata di aver rubato decine di milioni di profili e di averli utilizzati per costruire un potente algoritmo in grado di prevedere e influenzare le scelte di voto.

La somma dei "like" di Facebook di ogni singola persona profilata e dei test psicometrici ha permesso di segmentare in maniera più precisa la platea degli utenti.

In ambito politico, «segmentare» si traduce nel cogliere in maniera più precisa le tendenze ideologiche di un certo utente, calibrando i contenuti in maniera tale da solleticarne fragilità o motivi di tensione.

La strategia di Cambridge Analytica si è concentrata, in maniera più forte, sul blocco elettorale degli indecisi, ovvero quei cittadini che non sono radicati su una certa inclinazione ideologica, ma che possono essere convinti dagli argomenti più vicini alla propria sensibilità.

Quanto riportato da “Il Fatto Quotidiano”:

“Profilare e targettizzare: gli elettori di un determinato quartiere, area, contea o Stato vengono stratificati per censo, orientamento politico, propensione al voto, acquisti, comportamenti. Tutto è raccolto e organizzato in grandi database. I messaggi politici da sottoporre, a quel punto, se non personalizzati, sono altamente strutturati: nel tuo quartiere le fogne non vanno? Colpa del governatore democratico. Sono aumentati i furti? Lottiamo per il tuo diritto ad avere armi. E via così: alla rappresentanza sul territorio si sostituisce la politica on-demand. Decine di migliaia di messaggi diversi dicono una cosa sola: conosco il tuo problema e ti darò ciò di cui hai bisogno. Solo per il terzo dibattito con Hillary Clinton, lo staff di Trump ha testato 175 mila diversi messaggi.”¹

Nel 2014, Aleksandr Kogan chiese l'accesso alla banca dati di Kosinski per conto di un'agenzia di comunicazione specializzata in campagne elettorali, chiamata Strategic Communications Laboratories. Kosinski rifiutò Kogan, che perseguì il suo obiettivo, registrando una società di proprietà di SCL sotto il nome di Cambridge Analytica. Kogan, con essa, copiò il progetto di Kosinski: *My Personality Project*. Quest'ultimo, stava sperimentando l'applicazione del metodo OCEAN attraverso *Facebook* e i social network, sostenendo di poter delineare il profilo di un individuo attraverso la somma di un certo numero di “like” con un'accuratezza elevatissima. Tale studio verrà affrontato e approfondito nel prossimo capitolo.

¹ (Amato, 2018)

CAPITOLO 2

La profilazione

2.1 Tecniche di profilazione usate da CA

Cos'è la profilazione? Da *“Wikipedia, l'enciclopedia libera”*:

“Per **profilazione dell'utente** si intende correntemente l'insieme di attività di raccolta ed elaborazione dei dati inerenti agli utenti di servizi (pubblici o privati, richiesti o forzosi) per suddividere l'utenza in gruppi di comportamento.”

La vicenda ebbe inizio nel 2008, quando Michal Kosinski, psicologo del Centro Psicometrico (University of Cambridge), in collaborazione con David Stillwell e Thore Graepel, sperimentavano alcuni quiz che rispondessero a ciò che gli psicologi tutt'oggi chiamano "*Big Five*", i tratti della personalità, dall'inglese "OCEAN".

Cos'è il modello OCEAN?

Partiamo dal concetto di personalità. La personalità viene intesa come il risultato dell'interazione tra elementi biologicamente determinanti, fattori ambientali ed esperienze individuali.

Gli scienziati sostengono che essa possa essere decostruita in 5 tratti fondamentali: Openness to experience, Conscientiousness, Extraversion, Agreeableness, Neuroticism. La loro valutazione si basa su "IPIP-NEO", acronimo per "International Personality ITEM Pool - Neuroticism, Extraversion & Openness", questionario sulla personalità che valuta le persone in base al modello Big Five. È un test scientificamente basato su tali tratti ed è generalmente accettato in tutto il mondo come uno dei questionari più apprezzati e accurati.

È composto da 120 item, ognuno dei quali vengono valutati in base alla scala Likert a 5 punti, che varia da punto 1: "assolutamente vero per me" a 5: "assolutamente falso per me". Per esempio, come illustrato nella figura 2.1:

La scala Likert

Indichi il suo grado di accordo/disaccordo con le seguenti affermazioni

In completo disaccordo	In disaccordo	Né in accordo né in disaccordo	D'accordo	Completamente d'accordo
------------------------	---------------	--------------------------------	-----------	-------------------------

Figura 2.1: Esempio di scala Likert, <https://bit.ly/2KEkLS8>

Illustriamo i 5 tratti in maniera più approfondita, Figura 2.2:

1. **Openness to experience** (apertura mentale): “quanto si è aperti a nuove esperienze.” Un punteggio alto è indice di originalità, fantasia e ampi interessi. Una revisione degli studi, in contesti aziendali, ha rilevato che l’apertura mentale verso nuove esperienze può accuratamente predire chi diventerà e avrà successo come leader;
2. **Conscientiousness** (coscienziosità): “quanto si è perfezionisti.” Un punteggio alto indica l’ambizione, energia, perseveranza e accuratezza. Gli psicologi sostengono che la coscienziosità è il migliore elemento per predire sia il successo personale sia quello professionale. Infatti, le persone tendono ad essere fortemente d’accordo con asserzioni del tipo: “Sono sempre molto organizzato”;
3. **Extraversion** (estroversione): “quanto si è socievoli.” Questo è il tratto chiamato a volte “Surgency” (“surgency is a trait aspect of emotional reactivity in which a person tends towards high levels of positive affect”²). Un punteggio alto è indice di socialità, amore per il divertimento, affetto, essere amichevoli, loquaci e capaci di trarre energia dall’attività sociale;

²(Blandon, et al., 2010)

4. **Agreeableness** (amicalità): “quanto si è attenti e premurosi verso gli altri.” Un alto punteggio è indice di comprensione verso gli altri, generosità e affetto. La ricerca suggerisce che le persone piacevoli tendono ad essere più felici;
5. **Neuroticism** (stabilità emotiva): “quanto si è connessi alle proprie emozioni.” Un punteggio alto è indice di ansia, insicurezza e timidezza e le persone concordano solitamente con affermazioni del tipo: “Sono spesso ansioso o teso”.

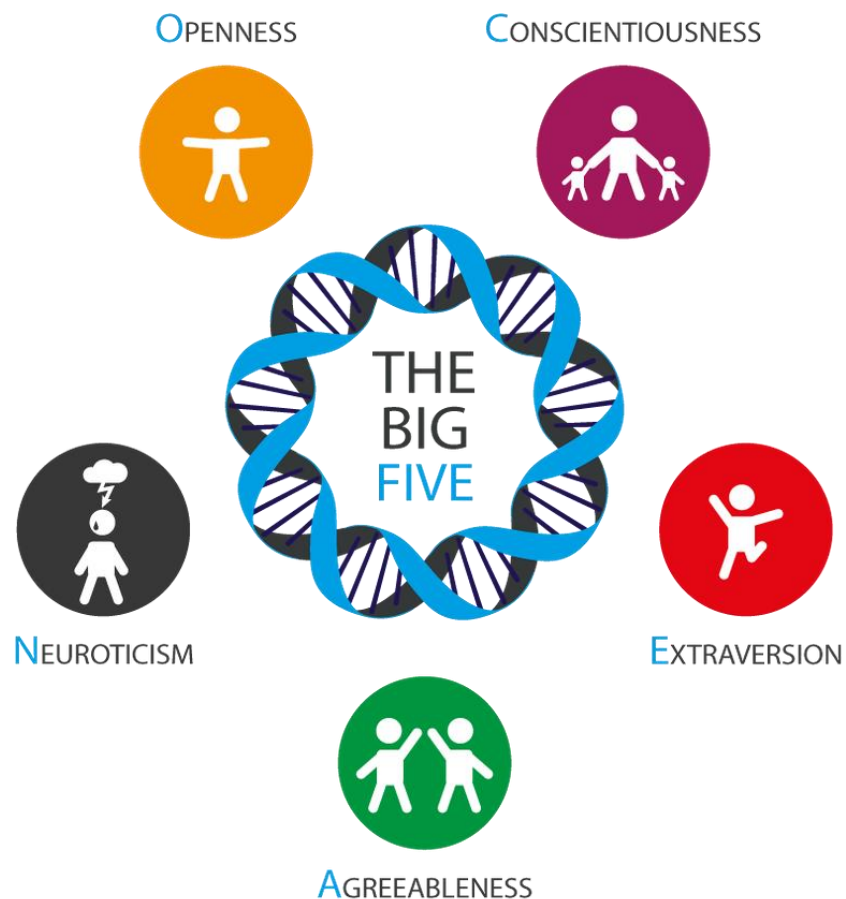


Figura 2.2: The Big Five, <https://bit.ly/2FTK53a>

Cos'ha a che fare tutto questo con le elezioni?

L'anno in cui la rivoluzione digitale si mostrò in tutta la sua potenza fu il 2013. Quando, con uno studio pubblicato sulla rivista Pnas, Kosinski mostrò la possibilità di predire caratteristiche sensibili di un utente basandosi su un certo numero di "like" di Facebook. Iniziò a correlare i dati derivanti dai test di personalità con i "MiPiace" del profilo social di una persona, dimostrando che, in media con 68 "like", era possibile predire il colore della pelle di un individuo con precisione del 95%, il suo orientamento sessuale (88%) e l'affiliazione politica al Partito democratico o repubblicano (85%). La ricerca non si fermò lì. Con una media di 170 "like" potevano essere determinati con grande precisione l'intelligenza, la religione, così come l'uso di alcool e sigarette. I social, dunque, si trasformarono immediatamente in un database in grado di fornire profilazioni perfette su elementi di incalcolabile delicatezza.

*"Non ho costruito una bomba", si giustificò Kosinski, "ma ho solo mostrato la sua esistenza."*³

Dopo pochi anni nel cratere di quell'esplosione si mosse Cambridge Analytica, che acquistando milioni di dati, è riuscita ad ottenere, secondo Alexander Nix, i profili di un numero enorme di elettori americani e a garantire ai suoi clienti la possibilità di inviare messaggi personalizzati a ognuno di loro basandosi su paure, bisogni e probabili comportamenti. Si tratta di dati di importanza fondamentale perché – spiegava Nix – “la personalità guida il comportamento, il comportamento influenza il voto”⁴. Infatti, attraverso i sofisticati algoritmi di CA non vennero prodotti profili “elettorali”, nel senso stretto del termine, delle persone, ma psicologici, emozionali e caratteriali, il che ha permesso una “targhettizzazione” estrema e chirurgica dei messaggi di propaganda. Ancora Nix, dichiara che “Cambridge Analytica ha portato Facebook a un nuovo livello: fino ad allora sono state organizzate campagne elettorali basate su concetti demografici, ovvero l'idea che tutte le donne dovessero ricevere lo stesso messaggio a causa del loro genere o che lo debbano ricevere gli afroamericani a causa della loro razza è un'idea veramente ridicola”⁵. Ciò che intendeva nel suo discorso era un concetto molto chiaro, Cambridge Analytica stava utilizzando la psicomatria, mai utilizzata finora.

³ (Kosinski, 2018) ; ^{4, 5} (Nix, 2017)

2.2 Teoria e fundamenta della profilazione attuata da Cambridge Analytica

Michal Kosinski, in collaborazione con David Stillwell e Thore Graepel, nel 2013 pubblicò sulla rivista scientifica statunitense *Pnas* uno studio chiamato “Private traits and attributes are predictable from digital records of human behavior” (I tratti e gli attributi personali sono prevedibili attraverso le impronte digitali del comportamento umano). Tale studio spiega come, analizzando post e “like” sul social network, si riesca a delineare un profilo completo della personalità dell’utente.

I “like” di Facebook rappresentano una classe generica di record e tracce digitali, simili alle query di ricerca di Google o acquisti online con carta di credito, che permettono di delineare i gusti e gli interessi di un soggetto. Lo studio si basa sulla raccolta di big data di ≈ 58.466 volontari negli Stati Uniti d’America, provenienti da varie fonti, tra le quali: la registrazione all’app “MyPersonality” tramite Facebook, un elenco di “like” personali ($n=170$ caduno) e infine i punteggi a test psicometrici (IPP). La Figura 2.3 ci illustra le fasi e l’evoluzione dello studio:

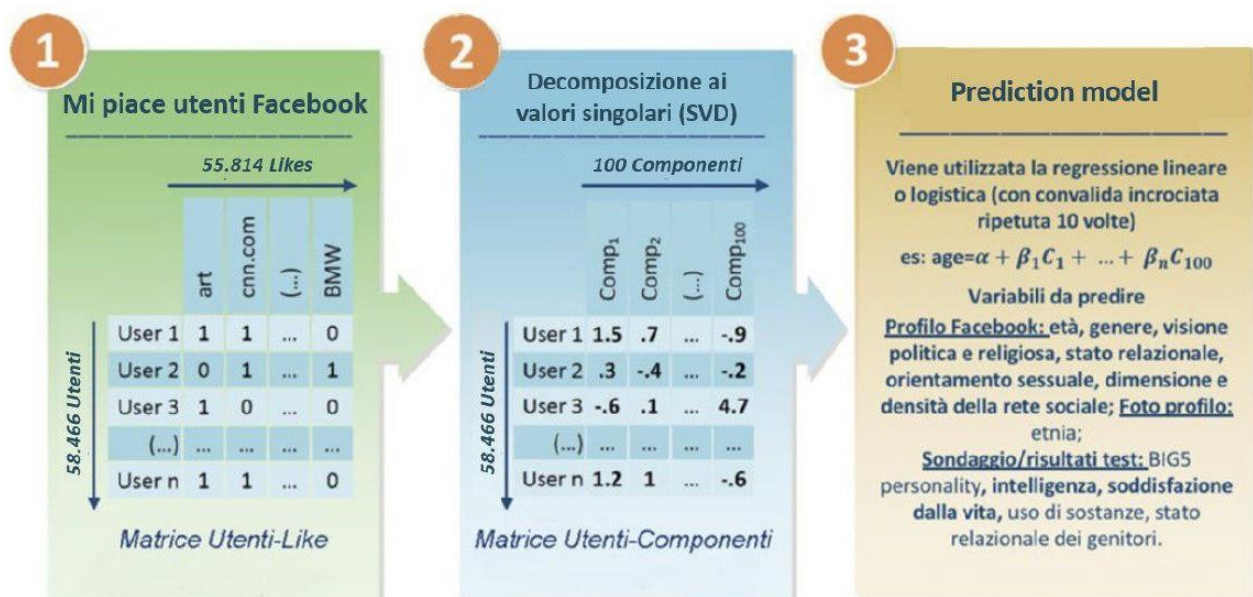


Figura 2.3: Prediction Model, <https://bit.ly/1mG7Go6>

- 1) Gli utenti e i loro corrispondenti like sono rappresentati in una matrice *User-Like*, le cui voci sono impostate su 1 se esiste un'associazione tra il like e l'utente, 0 altrimenti.
- 2) Nel passo successivo, la matrice viene ridotta utilizzando la "Singular Value Decomposition" (SVD), decomposizione di matrici basata sull'uso degli autovalori e degli autovettori, utilizzata nei problemi di mining per ridurre la dimensione dello spazio dei dati (a $k=100$) che stiamo trattando. Questa fattorizzazione può essere applicata a qualsiasi matrice rettangolare A avente dimensione (m,n) , scomponendola nel prodotto di tre matrici che presentano proprietà talvolta utili:

$$A = U\Sigma V^T \quad (1)$$

La matrice Σ è una matrice pseudo-diagonale avente dimensione (m,m) che contiene i valori singolari (detti $\sigma_1, \sigma_2, \dots, \sigma_i$ si calcolano a partire dalle radici degli autovalori della matrice $P=AAT$ e sono disposti diagonalmente in ordine decrescente) della matrice A ; Le matrici U e V^T sono matrici ortogonali unitarie e sono dette matrici dei vettori unitari, rispettivamente, destra e sinistra. I vettori colonna della matrice U corrispondono agli autovettori della matrice P . Quest'ultima è simmetrica e definita positiva, quindi i suoi autovalori sono non nulli e le radici degli autovalori sono reali. Di conseguenza i valori singolari sono definiti non nulli.

- 3) L'età o l'intelligenza, per esempio, sono variabili numeriche che vengono predette utilizzando un modello di regressione lineare, mentre le variabili dicotomiche, come l'orientamento sessuale o il genere, vengono pronosticate usufruendo della regressione logistica. La regressione formalizza e risolve il problema di una relazione funzionale tra variabili, misurate sulla base di dati campionari, estratti da un'ipotetica popolazione infinita. Le variabili, per poter essere inserite in un modello di regressione lineare multipla, devono essere del seguente tipo: variabile dipendente quantitativa (Y) e più variabili indipendenti quantitative o dicotomiche (X_1, X_2, \dots, X_p) come in (2):

$$Y \Leftarrow (X_1, X_2, X_3, X_4, \dots, X_p) \quad (2)$$

Si parla di iperpiano di regressione (3) perché sono presenti più di due variabili indipendenti (se fosse stata una, si sarebbe parlato di retta di regressione).

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_p + \varepsilon_i \quad (3)$$

I coefficienti di regressione parziali ($\beta_1, \beta_2, \dots, \beta_n$), presenti in (3), vengono stimati attraverso tale formula:

$$\beta = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{COVARIANZA(XY)}{VAR(X)} = \frac{\sigma_{XY}}{\sigma_X^2} \quad (4)$$

β rappresenta il cambiamento della variabile dipendente Y in seguito a una variazione unitaria della variabile indipendente X ed è data dal rapporto della covarianza (che fornisce la misura di quanto le due variabili varino assieme, ovvero della loro dipendenza) con la somma dei quadrati degli scarti dalla media (indice di dispersione dei dati). Per quanto riguarda α , è data dalla sottrazione:

$$\alpha = \bar{y} - \beta \bar{x} \quad (5)$$

con \bar{x} e \bar{y} rispettivamente media della variabile x_i e y_i .

Per quanto riguarda la regressione logistica, si tratta di un modello di regressione applicato nei casi in cui la variabile dipendente Y sia di tipo dicotomico, quindi riconducibile ai valori 0 e 1, come lo sono tutte le variabili booleane (vero o falso, maschio o femmina, vince o perde, sano o ammalato, ecc.).

Soffermandosi sulle variabili più difficili da predire, Kosinski ha utilizzato solo componenti principali SVD con $k=30$, in quanto, per esempio, per orientamento sessuale, consumo di sostanze stupefacenti o lo stato matrimoniale dei genitori si hanno meno informazioni, quindi un range limitato di dati degli utenti.

In ogni eventualità nel Prediction Model, viene applicata la **convalida incrociata o cross validation**, per 10 volte (tecnica statistica, usata in machine learning, che viene applicata in presenza di una buona numerosità del campione osservato).

Come si attua la convalida incrociata o cross validation?

Nella convalida incrociata, il campione originale viene ripartito casualmente in k sotto campioni di uguale dimensione ($k=10$ in questo caso). Un solo sotto campione, tra tutti, viene fissato e utilizzato come convalida per testare il modello, mentre i rimanenti $k-1$ sotto campioni vengono impiegati come dati di training o allenamento. Si avvia il processo di apprendimento induttivo sul training set residuo ($k-1/k$) per la costruzione dell'albero decisionale. Una volta costruito, si verifica l'efficacia predittiva dell'albero utilizzando la parte di convalida $1/k$ e si calcola il valore predittivo dell'albero parziale. Il processo di validazione incrociata viene poi ripetuto k volte e ognuno dei k sotto campioni viene utilizzato una sola volta come convalida. Si ottengono così k alberi decisionali parziali con ognuno un valore predittivo, che verranno usati per calcolare la media dei valori predittivi ottenuti nei k esperimenti. Si può così selezionare l'albero parziale con il valore predittivo più alto e utilizzarlo come modello. Il vantaggio di questo metodo è che tutti i campioni sono utilizzati sia per il training che per la convalida e ciascuno di essi viene impiegato per la convalida al più di una volta.

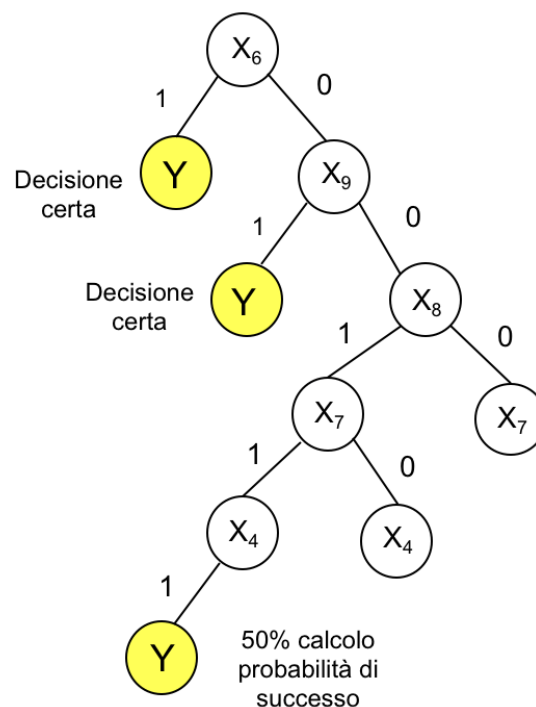


Grafico 2.1: albero decisionale parziale con valore predittivo

2.2.1 Predizione di variabili numeriche e dicotomiche

Nel caso in cui tra le due variabili non si possa ipotizzare un legame di dipendenza del tipo “causa-effetto”, viene applicato il metodo della correlazione lineare. Questo metodo consiste nel calcolare l'indice r , coefficiente di correlazione lineare di Pearson, che quantifica il *verso* (concorde o discorde) ed il *grado* della relazione tra le variabili. La Figura 2.4 rappresenta l'accuratezza della predizione delle variabili numeriche espressa attraverso il coefficiente di correlazione di Pearson, che oscilla tra valori effettivi e quelli previsti. Tale coefficiente r , che serve a misurare la correlazione tra le variabili fornite, è scaturito da:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}} = \frac{COVARIANZA(XY)}{\sqrt{VAR(X) VAR(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad (6)$$

Può assumere valori contenuti in un range che va da -1.00 (correlazione perfetta negativa) e $+ 1.00$ (correlazione perfetta positiva). Una correlazione uguale a 0 indica che non vi è alcuna relazione fra le due variabili.

La correlazione più alta è stata quella per la variabile età ($r = 0,75$), poi densità ($r = 0,52$) e numero di amici di Facebook ($r = 0,47$), seguiti dal tratto della personalità di apertura mentale ($r = 0,43$), estroversione ($r = 0,40$) e intelligenza ($r = 0,39$). I tratti psicologici fanno parte dei tratti latenti, ovvero quei tratti che non possono essere misurati direttamente ma solo in maniera approssimativa, valutando le risposte presenti nei questionari. Per questo motivo, nella Figura 2.6, sono presenti le barre trasparenti che indicano la precisione, sulla base di riferimento del questionario (punteggi previsti), espressa in termini di affidabilità test-retest (somministrazione dello stesso test allo stesso gruppo di soggetti, per due volte, dopo un certo intervallo di tempo). Il valore predetto più preciso è rappresentato dal tratto dell'*apertura mentale*, quello meno preciso, invece, è delineato da *satisfaction with life*, ovvero quel tratto che indica quanto si è soddisfatti della propria vita:

“The correlation between the predicted and actual Openness score ($r = 0.43$) was very close to the test–retest reliability for Openness ($r = 0.50$). This indicates that for the Openness trait, observation of the user’s Likes is roughly as informative as using their personality test score itself. For the remaining traits, prediction accuracies correspond to roughly half the questionnaire’s test-retest reliabilities. The relatively lower prediction accuracy for SWL ($r = 0.17$) may be attributable to the difficulty of separating long-term happiness from mood swings, which vary over time. Thus, although the SWL score includes variability attributable to mood, users’ Likes accrue over a longer period and, so, may be suitable only for predicting long-term happiness.”⁶

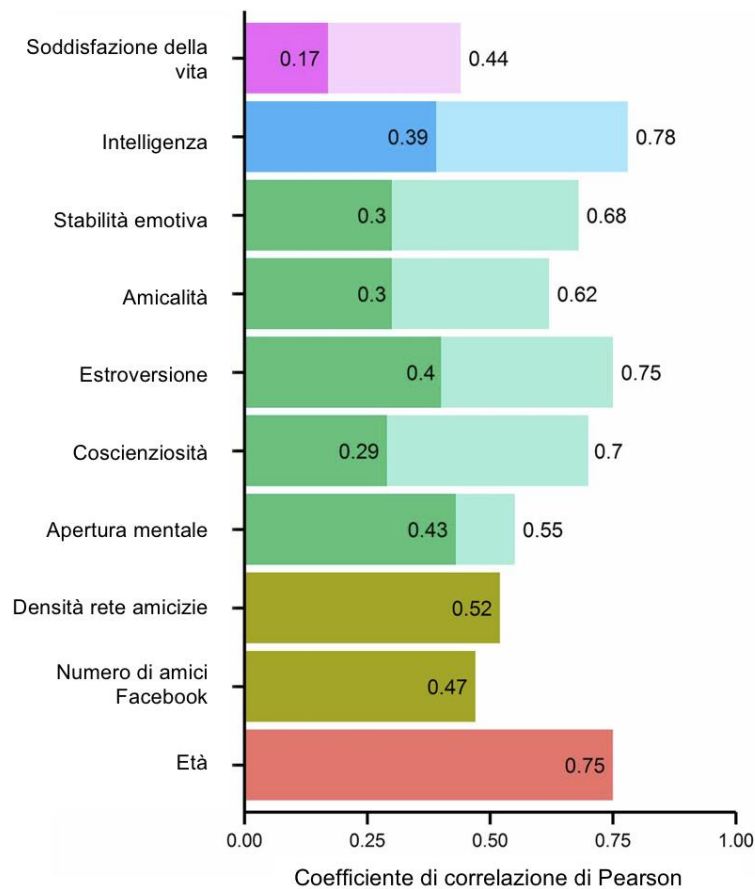


Figura 2.4: Predizione di variabili numeriche, <https://bit.ly/1mG7Go6>

⁶ (Michal Kosinski, 2013) : “La correlazione tra il punteggio di apertura previsto e quello reale ($r = 0,43$) era molto vicino all'affidabilità test-retest per Openness ($r = 0,50$). Questo indica che per il tratto Openness, l'osservazione dei Mi piace dell'utente è approssimativamente tanto informativa quanto l'utilizzo del punteggio del test di personalità stesso. Per i tratti rimanenti, l'accuratezza della previsione corrisponde a circa la metà delle affidabilità test-retest del questionario. L'accuratezza della previsione relativamente inferiore per SWL ($r = 0,17$) può essere attribuita alla difficoltà di separare la felicità a lungo termine dagli sbalzi d'umore, che variano nel tempo. Pertanto, sebbene il punteggio SWL includa la variabilità attribuibile all'umore, i Mi piace degli utenti maturano su un periodo più lungo e, quindi, possono essere adatti solo per prevedere la felicità a lungo termine.”

Per quanto riguarda le variabili dicotomiche, che possono assumere due soli valori, la precisione di previsione è data dall' *Area Under Curve* (AUC), che rappresenta la probabilità di classificare correttamente due utenti, scelti a caso, uno per classe. Il valore di AUC è compreso tra 0 e 1. La massima accuratezza è stata raggiunta, si può notare in Figura 2.5, per genere e etnia: maschi e femmine sono stati classificati correttamente nel 93% dei casi, invece afro-americani e caucasici-americani nel 95%.

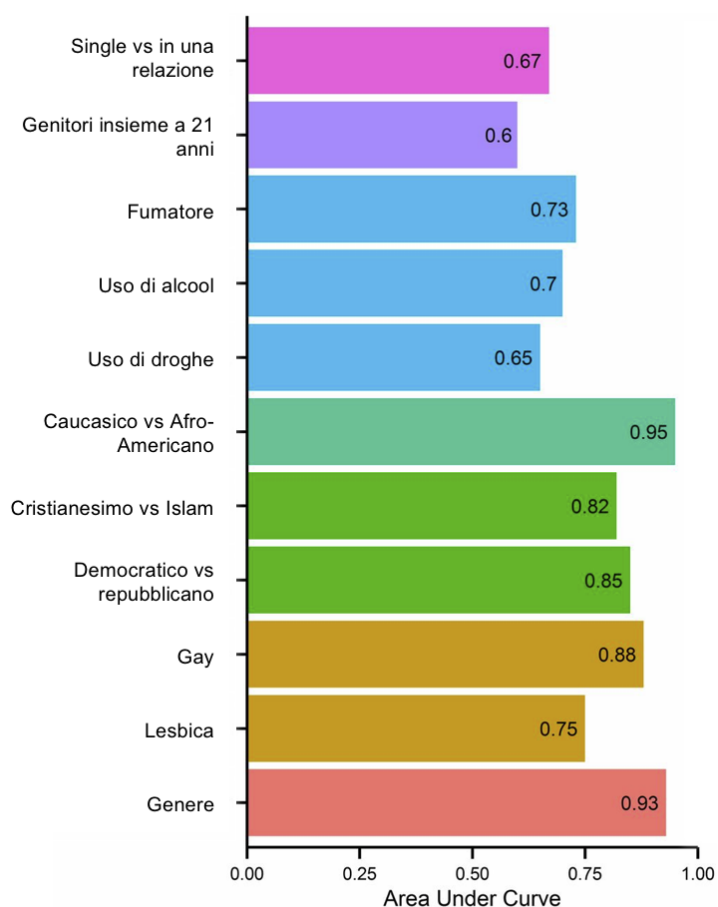


Figura 2.5: Predizione di variabili dicotomiche, <https://bit.ly/1mG7Go6>

Questo denota il fatto che esiste una grande differenza di comportamento online, espresso tramite i like, tra i due gruppi sia etnici che di genere. I risultati mostrati finora si basano su utenti Facebook con un range di likes che varia da 1 a 700. Il numero medio era di 68 per individuo, quanti più “mi piace” riusciamo ad ottenere e conoscere, maggiore sarà la precisione di previsione.

Si può notare in Figura 2.6, come un solo “like” per un dato utente può mostrare risultati con una precisione non trascurabile. La precisione è espressa dal rapporto tra AUC (genere) e il coefficiente di correlazione di Pearson (età e apertura mentale), al variare del numero di “Mi piace”. Circa il 50% degli utenti analizzati aveva almeno 100 likes, un 20% almeno 250.

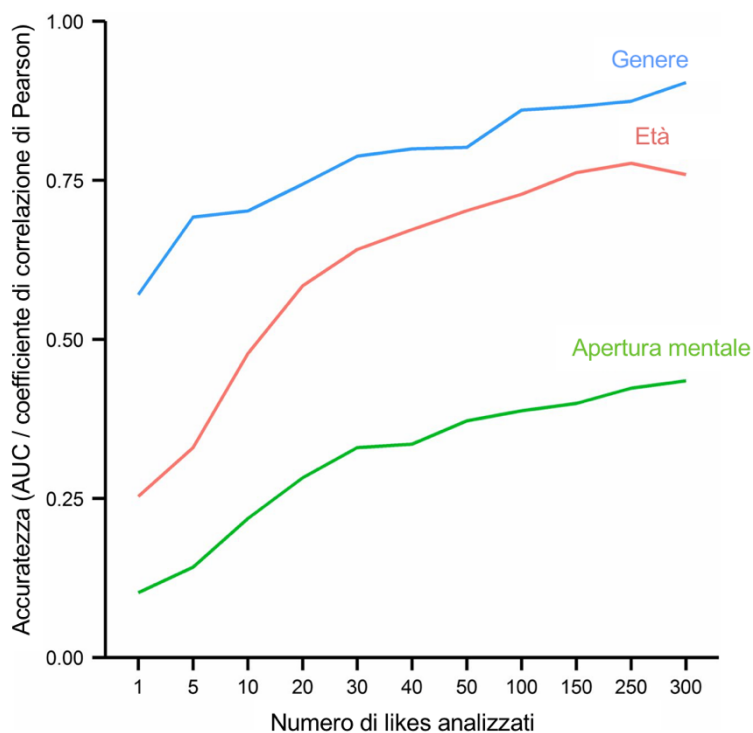


Figura 2.6: Precisione delle previsioni in funzione del numero di like disponibili,

<https://bit.ly/1mG7Go6>

Un "mi piace" alla cantante *Nicki Minaj*, per esempio, era fortemente correlato all'estroversione e un "mi piace" al personaggio *Hello Kitty* era associato all'apertura mentale.

In conclusione, abbiamo notato che si possono dedurre un'ampia varietà di attributi personali degli utenti, che vanno dall'orientamento sessuale al genere, all'intelligenza, in maniera automatica e precisa, utilizzando i like di Facebook e la psicometria, resa accessibile attraverso i questionari basati su i *Big Five*.

2.3 Software per la profilazione: Apply Magic Sauce

Gli scienziati dell'Università di Cambridge hanno mostrato che il social network di Mark Zuckerberg è in grado di attuare una profilazione accurata della nostra persona, grazie ad un algoritmo che sfrutta i nostri likes. La prova definitiva è *Apply Magic Sauce*, un software sviluppato dai ricercatori del centro di psicomatria dell'Università di Cambridge, tra i quali Michal Kosinski, che trasforma le nostre impronte digitali in un dettagliato profilo psicologico. Una sorta di identikit della nostra psiche, così come appare da ciò che siamo soliti condividere. David Stillwell, uno tra gli autori del software, incontrato nello studio precedentemente, in un'intervista al *Daily Mail*, disse: il problema vero «è che noi non riusciremo mai a vedere nel dettaglio cosa fa Facebook dietro le quinte con questi dati personali».

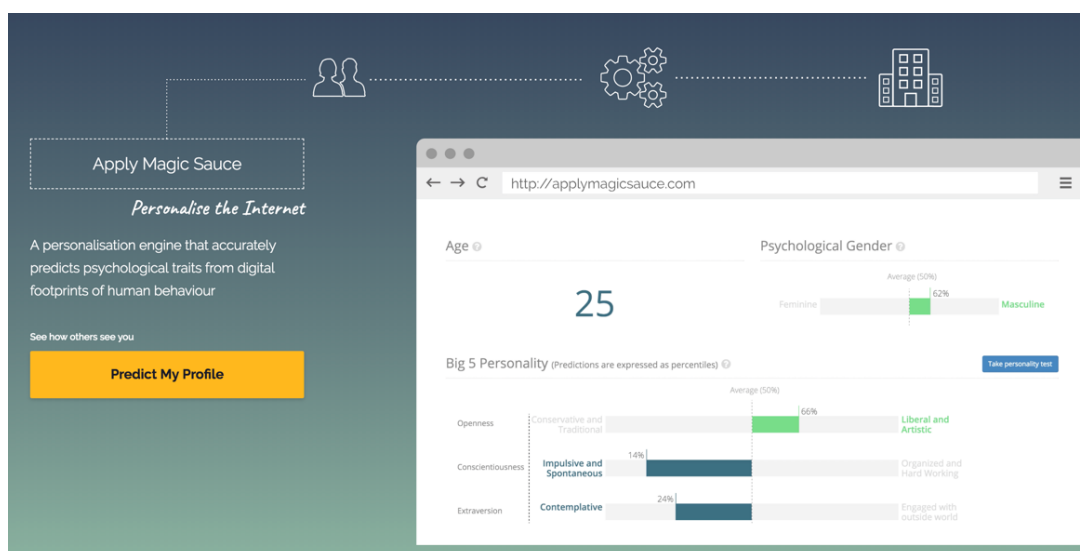


Figura 2.7: Interfaccia *Apply Magic Sauce*, <https://bit.ly/2qhGKKB>

Apply Magic Sauce, infatti, è la prima applicazione accessibile al pubblico dei non addetti ai lavori, la prima opportunità per metterci di fronte allo specchio e capire cosa i social fanno di noi.

Basta entrare nel sito: <https://applymagicsauce.com/>, accedere con le proprie credenziali Facebook, attendere alcuni istanti, affinché i dati vengano dati in pasto all'algoritmo, e alla fine potremmo conoscere la nostra età psicologica, il nostro vero orientamento sessuale, l'apertura mentale, l'estroversione, la stabilità emotiva, la situazione sentimentale, l'umore medio e il grado di intelligenza.

Il risultato ovviamente non è preciso al 100%, ma comunque molto accurato, nel fornire un quadro particolareggiato della propria immagine che si dà online. Per quanto riguarda la privacy, il team di Apply Magic Sauce ha assicurato che non copierà i dati e le informazioni del vostro profilo, né le ospiterà sul cloud dei propri server. Ecco alcuni esempi di risultati predetti da Apply Magic Sauce nelle Figure 2.8, 2.9, 2.10, 2.11:

Big 5 Personality (Predizioni sono espresse in percentuale)

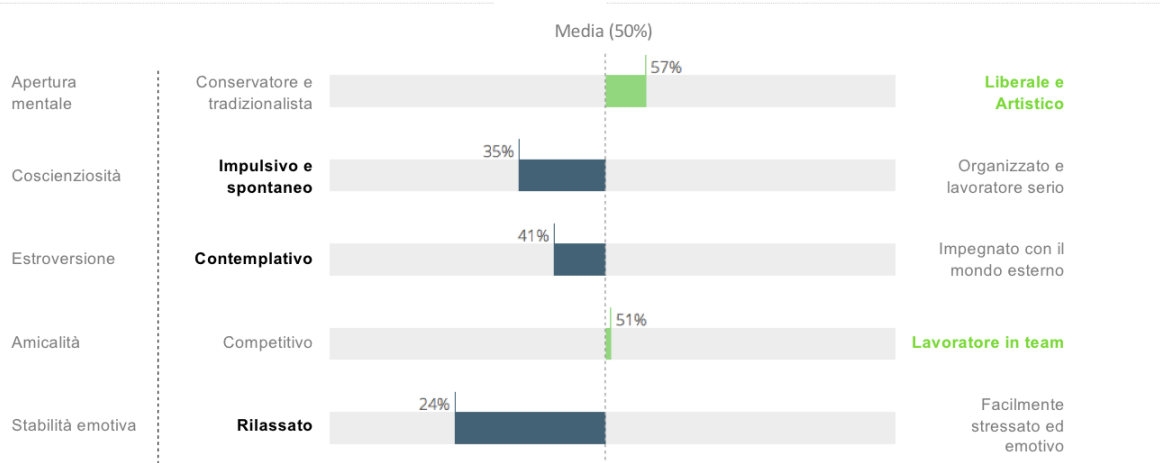


Figura 2.8: Risultati prodotti dal test della personalità (Big Five), <https://bit.ly/2qhGKKB>

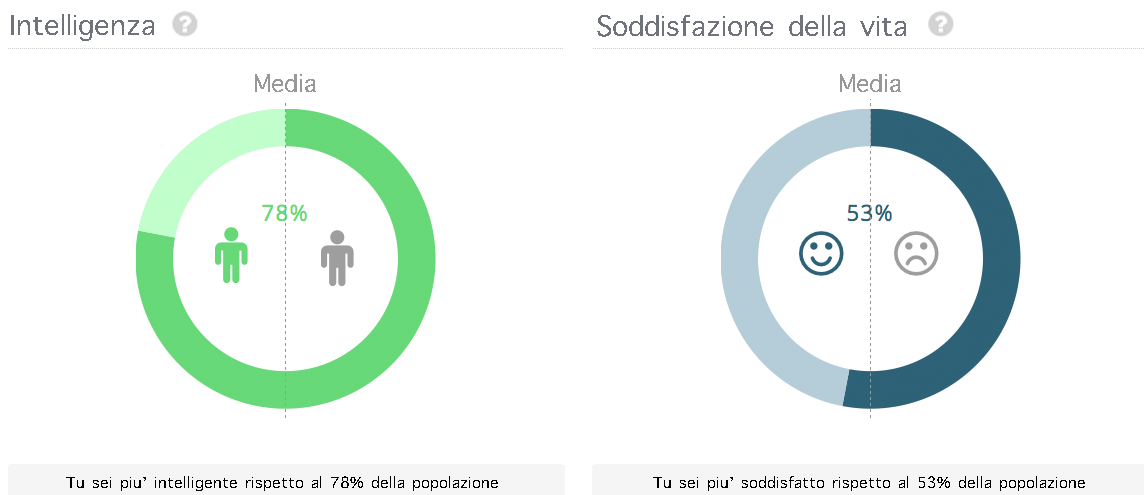
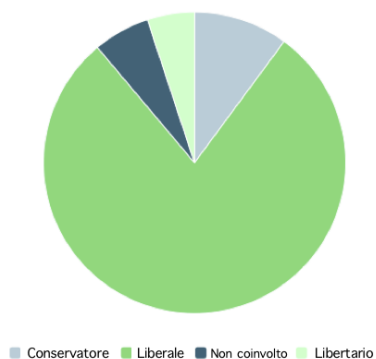


Figura 2.9: Risultati prodotti nell'ambito intelligenza e grado di soddisfazione nella vita, <https://bit.ly/2qhGKKB>

Orientamento politico



Orientamento religioso

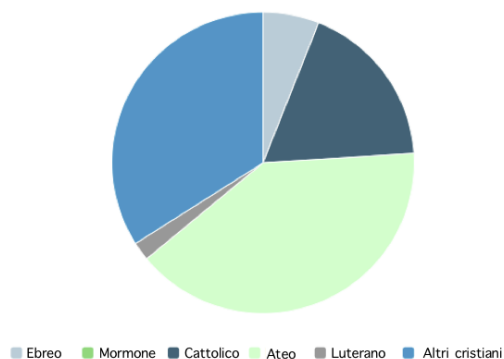
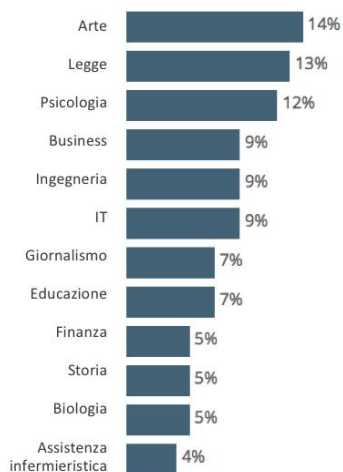


Figura 2.10: Risultati prodotti nell'ambito dell'orientamento politico e religioso, <https://bit.ly/2qhGKKB>

Formazione scolastica



Stato delle relazioni

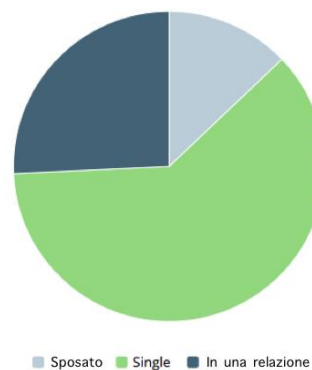


Figura 2.11: Risultati prodotti nell'ambito dell'educazione e delle relazioni personali, <https://bit.ly/2qhGKKB>

2.3.1 La Prediction API di Apply Magic Sauce

L'Università di Cambridge ha messo a disposizione la documentazione per integrare, ad una nostra ipotetica app, la Prediction API del software. Con API si intende *application programming interface*, ovvero ogni insieme di procedure disponibili al programmatore a formare un set di strumenti specifici per lo svolgimento di un determinato compito all'interno di un programma. Le API permettono di evitare ai programmatori di riscrivere ogni volta tutte le funzioni necessarie al programma dal nulla, permettendo il riutilizzo del codice. Le API funzionano utilizzando "richieste" e "risposte": quando un'API richiede informazioni da un'applicazione Web o un server Web, riceverà una risposta.

La Prediction API di Apply Magic Sauce è accessibile tramite la RESTful API (API che sfruttano al meglio la natura di HTTP e degli URI per fornire loro servizi): "<https://api.applymagicsauce.com/>". Per chiarezza, un'API RESTful fornisce: un URI base a cui accedere per ottenere i servizi, una sintassi degli URI delle entità interrogabili e modificabili, un media type attraverso cui ottenere e fornire dati da utilizzare nei servizi forniti (ad esempio XML, JSON, etc.) e una semantica associata all'uso dei vari verbi HTTP (GET, PUT, POST, DELETE).

Il luogo in cui un'API invia la richiesta e dove risiede la risorsa è chiamato endpoint, per esempio "<https://api.applymagicsauce.com/auth>" per l'autenticazione: questo restituirà, durante la registrazione, un token di autenticazione valido per chiamare, successivamente, il metodo di predizione basato per esempio sui likes di Facebook.

Come integrare la Prediction API ad un'applicazione Facebook:

Bisogna, innanzitutto, creare un'app Facebook e utilizzare il Facebook SDK per il linguaggio scelto (software development kit); raccogliere un numero rilevante di impronte digitali, per esempio "like" di Facebook; Integrare le funzioni al proprio progetto e utilizzare API RESTfull per attuare predizioni relative al tipo di impronta digitale che si possiede.

API RESTfull: Autenticazione, predizione della personalità attraverso i likes e attraverso il testo:

1) **Autenticazione:** per autenticarsi come utente è necessario passare alla funzione `def auth(customer_id, api_key)` i parametri `customer_id` e `api_key`, ottenuti durante la registrazione, loggarsi e, se la richiesta andrà a buon fine, verrà restituito un token di autenticazione (con validità temporanea) da usare per i metodi successivi. Per motivi di sicurezza, si effettua una chiamata POST nel metodo `requests.post()`, poiché i dati inviati non faranno parte dell'URL e soprattutto perché si passa l'`api_key`. I due argomenti passati sono URI e le credenziali dell'utente.

```
1. import json          # Python non supporta nativamente JSON, viene quindi fornito un pacchetto
2.                      # integrato per codifica e decodifica dei dati JSON
3. import requests     # requests è una libreria HTTP scritta in Python
4.
5.
6. def auth(customer_id, api_key):
7.     try:
8.         credentials = {
9.             'customer_id': customer_id,
10.            'api_key': api_key
11.        }
12.        response = requests.post('https://api.applymagicsauce.com/auth', json=credentials)
13.        response.raise_for_status()    # se abbiamo fatto una bad request, solleva l'eccezione
14.        return response.json()['token']
15.    except requests.exceptions.HTTPError as e:
16.        print e.response.json()
17.
18. #/auth
19. token = auth(1234, 'key')
```

© 2018 The Psychometrics Center

Possibili *status code* della risposta:

- **200:** OK, l'autenticazione è andata a buon fine;
- **400:** la richiesta non è valida, verrà stampato un messaggio;
- **403:** errore di autenticazione;
- **404:** controllare l'URI a cui si sta inviando la richiesta;
- **500:** API non risponde, potrebbe essere un errore temporaneo o essere in manutenzione.

2) Predizione della personalità attraverso i like di Facebook:

```
1) def predict_from_like_ids(token, like_ids):
2)     try:
3)         response = requests.post(url='https://api.applymagicsauce.com/like_ids',
4)                                 json=like_ids,
5)                                 headers={'X-Auth-Token': token})
6)         response.raise_for_status()
7)         if response.status_code == 204:
8)             raise ValueError('Not enough predictive like ids provided to make a prediction')
9)         else:
10)            return response.json()
11)    except requests.exceptions.HTTPError as e:
12)        print e.response.json()
13)    except ValueError as e:
14)        print e
15)
16) # /like ids
17) prediction_result = predict_from_like_ids(token, ["5845317146", "6460713406", "22404294985",
18)                                                  "35312278675", "105930651606", "171605907303",
19)                                                  "199592894970", "274598553922", "340368556015",
20)                                                  "100270610030980"])
```

© 2018 The Psychometrics Center

alla funzione `def predict_from_like_ids(token, like_ids)`, vengono passati come parametri il token di accesso, ottenuto dall'autenticazione, e la lista di `like_ids` contenuti in array JSON. Si possono aggiungere al metodo `requests.post()` parametri URI come:

- **Traits= trait1, trait2,...** : elenco di tratti della personalità che vogliamo prevedere, come per esempio “*Female*”, “*BIG5_Openess*”; consiste nell'applicare una limitazione a tutti i tratti che possono essere predetti, in quanto vengono restituiti tutti di default (se omesso traits);
- **Interpretations= true / false:** rappresentano interpretazioni, sempre facoltative, true se bisogna aggiungerle per attuare previsioni oppure false in caso contrario. Tale parametro non viene restituito di default;
- **Contributors = true / false:** true se la lista di `like_ids` influenza la previsione in maniera positiva, false altrimenti. Tale parametro non viene restituito di default.

Esempio di richiesta:

```
POST https://api.applymagicsauce.com/like_ids?traits=...
```

© 2018 The Psychometrics Center

Corpo di risposta previsto:

```
{
  "input_used": [number of like ids used in the prediction],
  "predictions": [
    {
      "trait": "[trait name]",
      "value": [predicted value, float]
    },
    ...
  ],
  "interpretations": [
    {
      "trait": "[trait name]",
      "value": [intepreted value, any]
    },
    ...
  ]
  "contributors": [
    {
      "trait": "[trait name]",
      "positive": [list of like ids that are influencing prediction in
        positive direction, strongest come first]
      "negative": [list of like ids that are influencing prediction in
        negative direction, strongest come first]
    },
    ...
  ]
} © 2018 The Psychometrics Center
```

Possibili *status code* della risposta:

- **200**: OK, la risposta dovrebbe contenere la previsione;
- **204**: non può essere fatta nessuna previsione in base agli id forniti;
- **400**: la richiesta non è valida, verrà stampato un messaggio;
- **403**: errore, problema riguardante il token;
- **404** controllare l'URI a cui si sta inviando la richiesta;
- **429**: il limite di utilizzo è stato superato per il tuo account (token scaduto);
- **500**: API non risponde, potrebbe essere un errore temporaneo o essere in manutenzione.

3) Predizione della personalità attraverso testo:

```
1) predict_from_text(token, text):
2)     try:
3)         response = requests.post(url='https://api.applymagicsauce.com/text',
4)                                 params={
5)                                     'source': 'OTHER'
6)                                 },
7)                                 data=text,
8)                                 headers={'X-Auth-Token': token})
9)         response.raise_for_status()
10)        return response.json()
11)    except requests.exceptions.HTTPError as e:
12)        print e.response.json()
13)
14) # /text
15) prediction_result = predict_from_text(token, 'Lorem ipsum dolor sit amet')
```

© 2018 The Psychometrics Center

Si tratta di previsioni riguardanti impronte digitali, quali il testo, che possiamo ritrovare in un sito web, in email, brochure, tweet, nel curriculum vitae o in altro. Si possono richiedere parametri URI, come nella previsione precedente (traits e interpretations), aggiungendo ad essi il parametro obbligatorio *source*. I risultati potrebbero dare output diversi se il parametro non è impostato correttamente:

- **Source = WEBSITE | EMAIL | BROCHURE | STATUS_UPDATE | TWEET | CV | OTHER:** riguarda il tipo di testo utilizzato per attuare la previsione.

Corpo di risposta previsto:

```
{
  "input_used": [currently it's a size of text sent to make a prediction],
  "predictions": [
    {
      "trait": "[trait name]",
      "value": [predicted value, float]
    },
    ...
  ],
  "interpretations": [
    {
      "trait": "[trait name]",
      "value": [intepreted value, any]
    },
    ...
  ]
} © 2018 The Psychometrics Center
```

Possibili *status code* della risposta:

- **200**: OK, la risposta dovrebbe contenere la previsione;
- **400**: la richiesta non è valida, verrà stampato un messaggio;
- **403**: errore, problema riguardante il token;
- **404**: controllare l'URL a cui si sta inviando la richiesta;
- **429**: il limite di utilizzo è stato superato per il tuo account (token scaduto);
- **500**: API non risponde, potrebbe essere un errore temporaneo o essere in manutenzione.

Allego la licenza MIT in Figura 2.14, licenza di software libero creata dal Massachusetts Institute of Technology, fornita da *The Psychometrics Centre*.

MIT License

Copyright (c) 2017 The Psychometrics Centre

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Figura 2.12: Licenza MIT per Apply Magic Sauce, <https://bit.ly/2uajjfc>

2.4 Privacy policy: trattamento dei dati personali

In questa vicenda, Facebook ha chiarito più volte che non si è trattato di una violazione dei dati personali intesa come *data breach*, ma di una violazione della privacy policy, in quanto non è consentito da parte di sviluppatori (Kogan) la comunicazione a terzi dei dati così raccolti (da sottolineare che Kogan necessitava di questi dati per “fini accademici”).

Tra il 2007 e il 2014 le norme di Facebook hanno consentito la raccolta di informazioni private sui propri utenti, (erano ancora visibili i mi piace e commenti di post altrui sul proprio profilo) attraverso app di terze parti tramite l'API del social. Una volta modificata l'API nel 2015 la condivisione dei dati risultò più restrittiva: Facebook non consentiva più alle app di attingere ai dati tramite il solo utilizzo, ma bisognava acconsentire esplicitamente. Tuttavia, l'app “thisisyourdigitallife” ebbe il consenso da parte degli utenti, ma non da parte degli amici degli utenti. Facebook ha dichiarato, nella sua privacy policy, che le informazioni degli amici possono essere utilizzate solo per migliorare l'*user experience* dell'utente ed è assolutamente vietata la vendita o l'utilizzo di tali informazioni per pubblicità (cosa che ha fatto CA per la controversa campagna presidenziale di Trump, per le elezioni locali negli Stati Uniti, per il referendum sulla Brexit e molte altre campagne elettorali a livello globale).

Ancor più grave la circostanza, evidenziata da un dipendente di CA, che Facebook era già a conoscenza da ben due anni della violazione della privacy policy e che in tutto questo periodo non abbia adottato alcuna misura nei confronti di detta società (né abbia informato i propri utenti).

Dal punto di vista strettamente giuridico, quindi, ciò significa che il Titolare, venuto a conoscenza della violazione dei diritti degli interessati da parte di un terzo, in violazione degli accordi intercorsi tra il medesimo Titolare ed il terzo, potrebbe aver deciso di non far nulla. D'altra parte, la comunicazione a terzi non autorizzata dei dati personali rappresenta una violazione di confidenzialità degli stessi, pur non derivante da un incidente di sicurezza, ma come tale potrebbe essere trattata in maniera analoga ad un incidente di sicurezza.

Secondo l'art. 4, comma 1, n. 12 del GDPR per violazione di dati personali deve intendersi *“la violazione di sicurezza che comporta accidentalmente o in modo illecito la distruzione, la perdita, la modifica, la divulgazione non autorizzata o l'accesso ai dati personali trasmessi, conservati o comunque trattati”*⁷.

⁷ (PARLAMENTO EUROPEO, CONSIGLIO DELL'UNIONE EUROPEA, 2016)

CAPITOLO 3:

Microtargeting politico attuato da CA

3.1 Facebook come strumento d'influenza politica

Da "Wikipedia, l'enciclopedia libera":

“Il **microtargeting**, spesso utilizzato dai partiti politici e dalle campagne elettorali, include tecniche di data mining di marketing diretto che prevedono una segmentazione predittiva del mercato (ovvero l'analisi dei cluster). È utilizzato dai partiti politici repubblicani e democratici degli Stati Uniti, così come i candidati per monitorare i singoli elettori e identificare potenziali sostenitori.

Le tattiche di microtargeting si basano sulla trasmissione di un messaggio personalizzato a un sottogruppo dell'elettorato sulla base di informazioni uniche su quel sottogruppo.”

Il *microtargeting psicografico* attuato da Cambridge Analytica è una tipologia di marketing politico innovativo, fondato sulla misurazione della personalità degli elettori in base alle loro tracce digitali e nella pratica di influenza attraverso l'invio di messaggi personalizzati.

Cambridge Analytica, dopo aver profilato milioni di cittadini americani, è passata a tale fase, inviando agli utenti Facebook annunci pubblicitari sulla base della personalità dedotta dal modello e adattando i propri annunci a persone con caratteristiche particolari. Inoltre, Cambridge Analytica, grazie alla funzione "Lookalike" di Facebook (che trova utenti ritenuti simili a un determinato gruppo), avrebbe potuto anche rivolgersi a persone che non aveva profilato.

Nix ha mostrato come sia possibile rivolgersi in modo differenziato agli elettori di ogni categoria psicografica, prendendo come esempio il secondo emendamento della costituzione degli Stati Uniti, che garantisce a ogni cittadino il diritto di possedere armi da fuoco: “Per convincere le persone fortemente nevrotiche e coscientose, serve la minaccia del furto in casa e la sicurezza rappresentata da

un'arma"⁸. Come si nota in Figura 3.1, l'immagine a sinistra dello schermo raffigurava la mano di un intruso che sfonda una finestra. L'immagine sulla destra ritraeva un uomo e un bambino in piedi in un campo al tramonto: entrambi impugnavano armi da fuoco e stavano chiaramente sparando alle anatre. "Questa invece serve per i soggetti chiusi e disponibili, quelli che hanno a cuore le tradizioni, le abitudini e la famiglia"⁹.



Figura 3.1: Alexander Nix al Concordia Annual Summit (New York)

Messaggi customizzati basati sui profili psicografici, <https://bit.ly/2qD8FbL>

“Praticamente ogni messaggio lanciato da Trump si basava su dati digitali”, ha ricordato Nix. Il giorno del terzo dibattito televisivo fra Trump e Clinton, la squadra del candidato repubblicano ha testato, soprattutto attraverso Facebook, 175mila variazioni di inserzioni sui temi della campagna elettorale. Nella maggior parte dei casi i messaggi differivano tra loro solo per dettagli microscopici, con l’obiettivo di rivolgersi ai destinatari nel modo più consono al loro profilo psicologico. C’erano titoli diversi, colori e didascalie diversi, accompagnate da foto o video. Come lo stesso Nix ha spiegato, in un’intervista a *Das Magazin*, queste variazioni quasi impercettibili servono a raggiungere anche i gruppi più piccoli: “In questo modo

⁸, ⁹ (Nix, 2017)

siamo in grado di rivolgerci in modo mirato ad un intero villaggio come ad un condominio e perfino a singole persone”.

A Miami è presente un quartiere chiamato *Little Haiti*. Per evitare che i suoi abitanti votassero per Clinton, lo staff che seguiva la campagna elettorale di Trump ha messo in circolazione la notizia del fallimento della *Clinton foundation* in seguito al terremoto di Haiti: stavano cercando di tenere lontani dai seggi i potenziali elettori della candidata democratica. L'obiettivo è stato raggiunto attraverso i cosiddetti *dark post*, cioè inserzioni sponsorizzate che si presentano come “ultimissime notizie”. I *dark post* sono comparsi su Facebook e possono essere visti solo dagli utenti che hanno profili specifici. Un esempio sono i video rivolti agli afroamericani, in cui Hillary Clinton definiva “predatori” i maschi neri.

Le indagini condotte fino a oggi (la tv *Channel 4* ha ripreso in segreto il capo della società, Alexander Nix, che ammette di aver usato trappole e metodi illegali per influenzare elezioni straniere) hanno accertato che nel corso della campagna elettorale pro-Trump furono utilizzati numerosi account fasulli e *bot* per diffondere notizie false e altri contenuti finalizzati a screditare Hillary Clinton.

Ogni giorno venivano pubblicati decine di migliaia di post, soprattutto in occasione dei dibattiti tv e di altri appuntamenti elettorali: l'efficacia dei post veniva analizzata in tempo reale, così da potere privilegiare quelli che maggiormente erano in grado di influenzare le opinioni dell'elettorato.

In ogni caso la Cambridge Analytica ha suddiviso la popolazione statunitense in 32 tipi di personalità e ha concentrato i suoi sforzi solo su 17 stati.

Inoltre, così come Kosinski era arrivato alla conclusione che gli uomini che mettono like su Facebook ai cosmetici Mac hanno qualche probabilità in più di essere gay, la Cambridge Analytica ha scoperto che la preferenza per le automobili di fabbricazione statunitense era tipica dei potenziali elettori di Trump. La decisione di concentrarsi sul Michigan e sul Wisconsin nelle settimane conclusive della campagna presidenziale è stata presa sulla base dell'analisi dei dati. Insomma, il candidato Trump è diventato lo strumento per l'applicazione di un modello.

L'ex dipendente di CA e informatore Christopher Wylie ha spiegato a riguardo:

"Invece di stare in piedi nella piazza pubblica e dire quello che pensi, stai sussurrando all'orecchio di ogni singolo elettore. E potresti sussurrare una cosa ad un elettore ed un'altra cosa ad un altro elettore ".¹⁰

Facebook si è dimostrato l'arma più efficace per la vittoria di Trump, consolidata dal fatto che ha investito molto di più nella campagna digitale che in quella televisiva.

¹⁰ (Carole Cadwalladr, 2018)

3.2 Analisi scientifica di Kosinski: “Targeting psicologico come approccio efficace alla persuasione digitale di massa”

Michal Kosinski, nel 2017, insieme a Sandra Matz, ricercatrice della *Columbia Business School* e della *Gideon Nave della Wharton School of Business*, hanno condotto una serie di test che dimostrano l’efficacia del microtargeting basato sulla personalità. Questo studio risponde alla domanda: “Il targeting psicologico è uno strumento efficace di propaganda digitale?”. I ricercatori confermano che si tratti di un efficace strumento di persuasione digitale di massa.

I ricercatori hanno condotto tre campagne pubblicitarie sperimentali su Facebook. Il primo studio riguarda una campagna che vuole promuovere una linea di cosmetici mirata a donne con tratti psicologici, elevati e non, di estroversione e apertura all’esperienza. L’annuncio per estroverse raffigura, come in Figura 3.2, una donna che balla con slogan: “Balla come se nessuno guardasse.” Al contrario, la pubblicità per introversive ritrae una donna che contempla il suo viso in uno specchio con slogan silenzioso: “La bellezza non deve gridare.”



Figura 3.2: Campagna pubblicitaria su una linea di cosmetici,

<https://bit.ly/2je2cUF>

La pubblicità ha raggiunto $\cong 3$ milioni di utenti, ha attirato $\cong 10$ mila clic e ha portato a 390 acquisti sul sito del rivenditore di cosmetici. Il genere femminile si è dimostrato più propenso all’acquisto dopo aver visto un annuncio che corrispondesse alla propria personalità. È stata condotta un’analisi di regressione logistica, come in

Figura 3.3, per tasso di conversione (metrica di marketing digitale che riflette il numero di conversioni, come download di app o acquisti di negozi online, in relazione al numero di volte in cui l'annuncio è stato mostrato) con avvenuta conversione =1, nessuna conversione=0, usando i tratti personali del pubblico, la personalizzazione dell'annuncio e la loro interazione bidirezionale come predittori. Nell'ambito del marketing, si utilizza la regressione logistica per evidenziare l'incisività delle campagne pubblicitarie sulla propensione all'acquisto del consumatore (il cliente acquista (Y=1) o non acquista (Y=0) un dato prodotto in seguito ad una campagna promozionale).

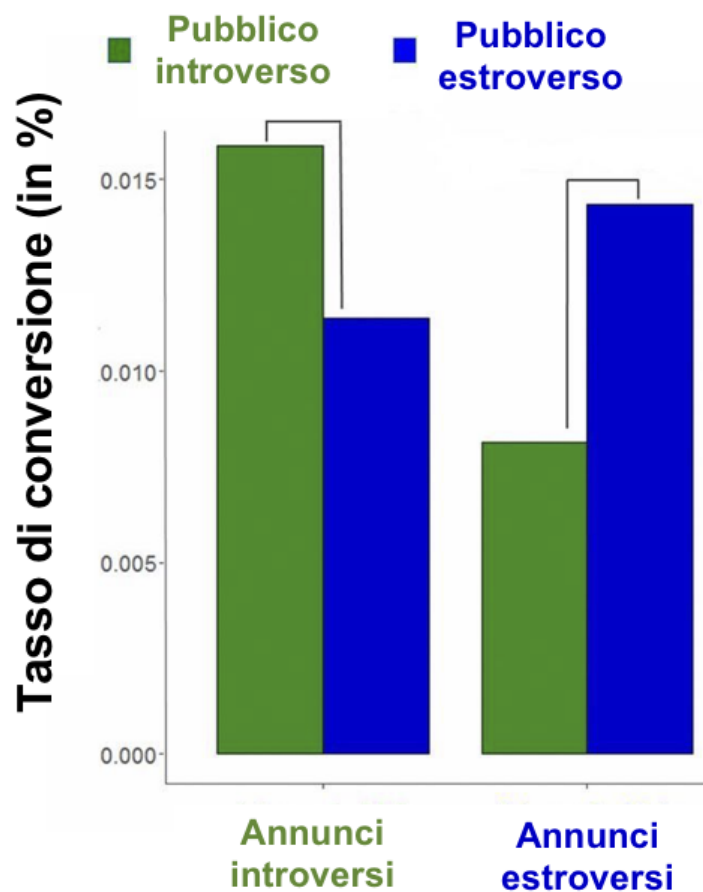


Figura 3.3: Effetti di interazione del pubblico con l'annuncio personalizzato, basato su i tratti di personalità introversi e estroversi, riguardante i tassi di conversione, <https://bit.ly/2je2cUF>

Il secondo studio, simile al primo, riguarda una campagna pubblicitaria che ha per oggetto messaggi persuasivi per incentivare il download di un'applicazione di cruciverba per soggetti con alta e bassa apertura mentale alle novità (high openness/low openness).

L'inserzione ha raggiunto $\cong 84$ mila utenti, ha attirato $\cong 1.200$ clic e ha generato 500 download dell'applicazione. Le persone che erano state identificate come molto aperte sono state incuriosite dall'annuncio che ha per slogan: "scatena la tua creatività" su "un numero illimitato" di enigmi. Al contrario, come raffigurato in Figura 3.4, l'inserzione opposta proponeva: "il cruciverba che ha sfidato i giocatori per generazioni".



Figura 3.4: Campagna pubblicitaria per un'applicazione di cruciverba per smartphone, <https://bit.ly/2je2cUF>

Le persone low openness hanno installato l'applicazione con più frequenza, invece le restanti persone (high openness) non hanno mostrato differenze di conversione significative, come raffigurato in Figura 3.5.

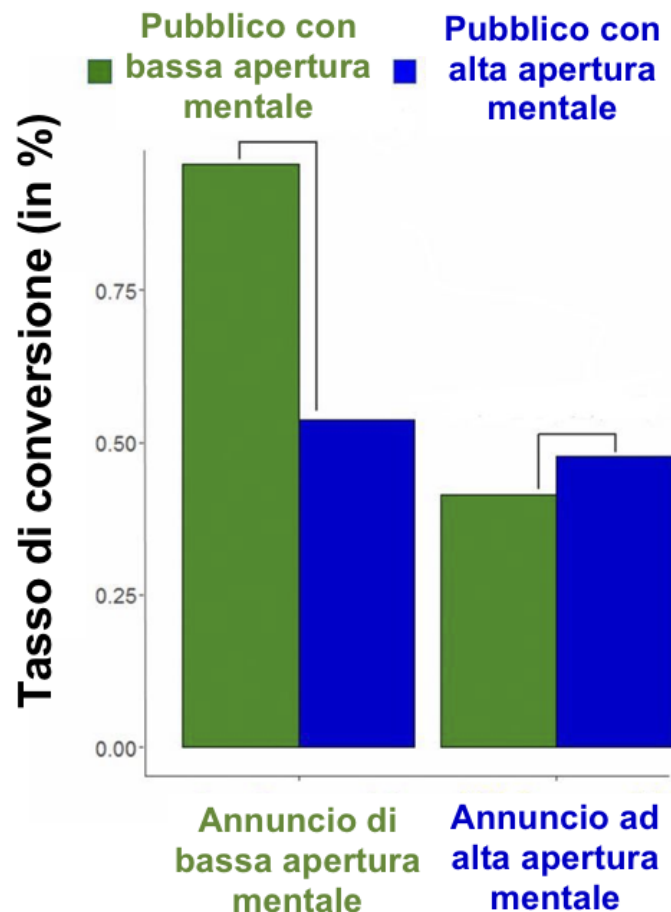


Figura 3.5: Effetti di interazione del pubblico con l'annuncio personalizzato, basato su i tratti di personalità di apertura mentale, riguardante i tassi di conversione, <https://bit.ly/2je2cUF>

Nel terzo e ultimo test, Kosinski e i suoi colleghi hanno testato pubblicità rivali per un videogioco, fortemente pensato per introversi. Il primo annuncio presentava uno slogan standard pieno di azione: "Pronto? Fuoco! Prendi ora l'ultimo rompicapo!". Il secondo annuncio, mirato alla categoria di introversi, diceva: " Phew! Giornata faticosa? Che ne dici di un puzzle da abbattere?". In questo caso, gli annunci per introversi hanno generato il 30% e il 20% in più, rispettivamente di clic e download. Hanno dimostrato che questo tipo di marketing è in grado di attirare fino al 63% di interazioni in più nelle campagne su Facebook e anche 1.400 conversioni in più, proponendo prodotti e messaggi confezionati su misura per la personalità di ogni consumatore.

Kosinski, al termine di questo studio, afferma che è probabilmente impossibile proibire il targeting psicologico come strumento di propaganda politica, ma afferma che le persone possono prendere coscienza di come si metta in atto. Possono anche essere in grado di attuare politiche che impediscano abusi.

"È un po' come il fuoco", dice "puoi usare il fuoco per riscaldare la tua casa e bruciarla. Ma non puoi vietare il fuoco ed impedire ad alcune persone di commettere un incendio doloso. Quello di cui hai bisogno sono i vigili del fuoco e le attrezzature antincendio."

3.3 Tecnologie usate da Cambridge Analytica per influenzare il popolo americano targettizzato

L'assetto tecnologico di Cambridge Analytica, per quanto riguarda *l'influenza*, comprende l'utilizzo dei **cookies**, **Facebook Audience Insight** e **Facebook Lookalike**.

- Un ruolo determinante in questo processo l'hanno avuto i *cookies*, file inviati sul nostro computer, capaci di tracciare i nostri comportamenti di navigazione online. Negli ultimi anni il loro uso si è esteso ulteriormente perché si sono trovate tecniche in grado di trasformarli in strumenti di marketing e di profilazione a fini pubblicitari. Infatti, “attraverso i cookie le piattaforme automatizzate per l'acquisto degli spazi pubblicitari riescono ad intercettare uno specifico segmento di elettori/consumatori, mentre naviga in rete, mostrando così il corretto messaggio al giusto consumatore”¹¹, i cosiddetti *cookie di terze parti*.
- “Il modello predittivo di Cambridge Analytica” venne “potenziato con lo strumento di analisi e scelta delle audience del social network di Zuckerberg, *il Facebook Audience Insight*. Si tratta di una potente funzionalità delle Facebook Ads, che permette di analizzare un pubblico creato manualmente (pubblico personalizzato) o definito attraverso interessi, comportamenti, aspetti demografici e connessioni.

Non appena si accede al tool, si visualizza un popup che chiede di scegliere un pubblico di partenza, come in Figura 3.6.

Definendo i parametri dal pannello di controllo sulla sinistra si noteranno i dati aggregati sulla destra. Ogni dato è riportato in percentuale e confrontato con la % dell'intero pubblico di utenti Facebook. In particolare, sono presenti “parametri socio-demografici come l'età, il sesso, la localizzazione geografica, il lavoro, il titolo di studio, che sono incrociati con i dati dei comportamenti degli utenti su Facebook e nel mondo reale”¹² poi, mi piace

¹¹, ¹² (Sisti, 2017)

a pagine (le categorie di pagine con più seguaci sul target analizzato e le pagine facebook che potrebbero essere pertinenti per il pubblico selezionato), luogo (stati, città e lingue di visualizzazione), attività (azioni svolte su facebook come commenti, “mi piace” cliccati, condivisioni) e dispositivi utilizzati.

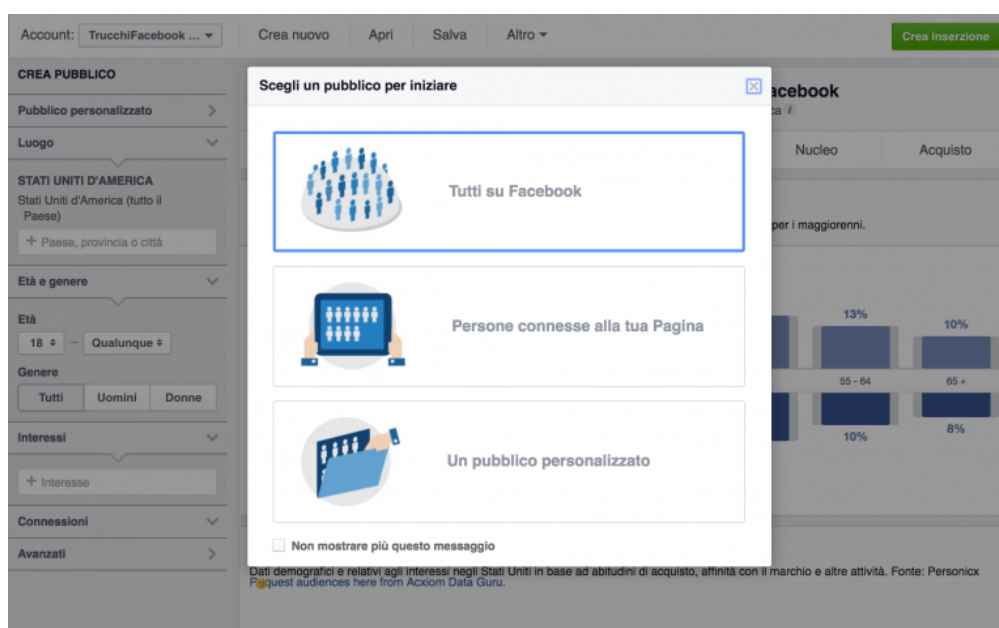


Figura 3.6: Scelta pubblico da monitorare, <https://bit.ly/1mGh8sF>

“Interessi, attività, like evidenziati dai comportamenti di navigazione sulle pagine del social network vengono aggiunti nell’analisi,” – come si può notare in Figura 3.7 - “consentendo l’invio di *specifici messaggi pubblicitari* a ben identificati segmenti di utenti (*micro-targeting*). Le infinite potenzialità di targeting e personalizzazione disponibili negli strumenti pubblicitari del social network, spiegano la crescita impetuosa del social media advertising.”¹³

Si tratta di uno strumento da utilizzare in molteplici casi, sia per effettuare valutazioni d’ingresso in un mercato nuovo, sia subito dopo, per valutare i risultati di una campagna pubblicitaria. In seguito alla profilazione da parte di Cambridge Analytica, è stata rimossa l’opzione di scelta del pubblico personalizzato, lasciando inalterati gli altri parametri.

¹³ (Sisti, 2017)

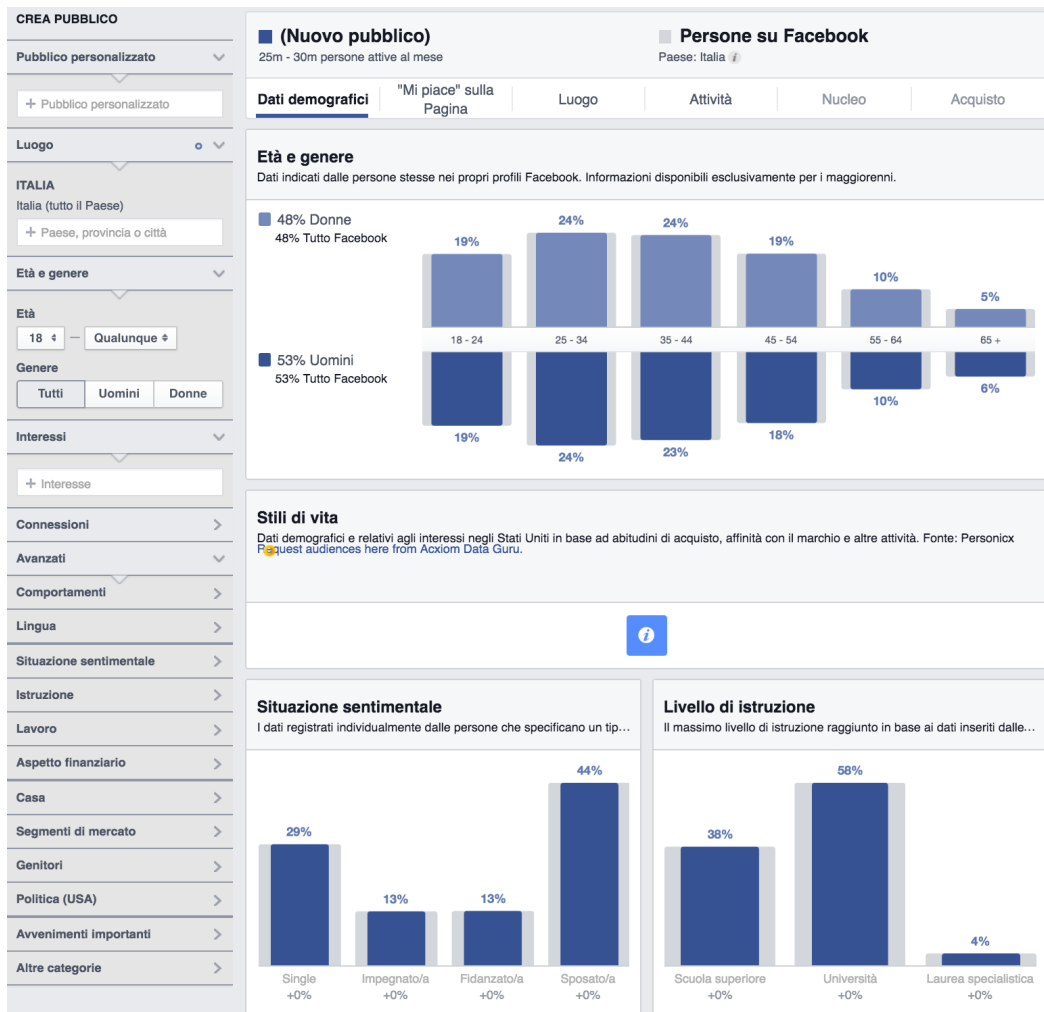


Figura 3.7: Facebook Audience Insight, <https://bit.ly/1mGh8sF>

- I dati ottenuti dagli utenti profilati sono stati impiegati anche per puntare al "Lookalike Audiences" su Facebook (pubblico simile). "Il pubblico simile" ha consentito a CA di trovare più persone simili agli utenti profilati da influenzare, (in base alla personalità) e ciò ha reso possibile modulare il contenuto dei messaggi per renderlo più efficace e persuasivo. Per iniziare e sfruttare lookalike audiences è necessario creare/usare dei pubblici personalizzati (Figura 3.8): si possono scegliere le dimensioni di tale pubblico, Facebook consiglia di usare un gruppo di utenti di almeno 100 persone per poter creare un pubblico estremamente affine. Maggiore è il numero del pubblico personalizzato, più simili saranno le persone trovate.

Una volta individuato (la creazione del pubblico simile impiega dalle 6 alle 24h), si crea l'inserzione personalizzata e questa verrà visualizzata agli utenti di destinazione.



Pubblico chiave

Seleziona il pubblico manualmente in base a caratteristiche come età e luogo.



Pubblico personalizzato

Carica la tua lista di contatti per connetterti con i tuoi clienti su Facebook.



Pubblico simile

Usa le informazioni sui tuoi clienti per trovare persone simili a loro su Facebook.

Figura 3.8: Facebook Lookalike, <https://bit.ly/1P3mYiD>

CAPITOLO 4

Sviluppi recenti

4.1 Innovazioni tecnologiche e cambiamenti in seguito allo scandalo di Cambridge Analytica

- 1) **Chiusura Cambridge Analytica:** la chiusura di Cambridge è stata dichiarata ufficialmente il 2 maggio 2018. Ha annunciato lo stop "immediato" di tutte le sue attività e l'avvio delle procedure di insolvenza in Gran Bretagna. Tutti i dipendenti hanno ricevuto l'ordine di restituire qualunque tecnologia e computer che sia in loro possesso. Cambridge aveva ottenuto contratti per 15 milioni di dollari durante la campagna elettorale del 2016 negli Stati Uniti, che si è rivelata allo stesso tempo, l'apice del suo business e del suo successo e l'inizio della sua precipitosa caduta. La decisione a sorpresa segue la perdita di numerosi clienti dopo lo scandalo dei dati di Facebook e le potenziali elevate spese legali delle indagini avviate per far luce sull'accaduto. Nell'annunciare la cessazione delle attività, Cambridge Analytica ribadisce la propria posizione e si difende:

"Over the past several months, Cambridge Analytica has been the subject of numerous unfounded accusations and, despite the Company's efforts to correct the record, has been vilified for activities that are not only legal, but also widely accepted as a standard component of online advertising in both the political and commercial arenas."¹⁴

Inoltre, saranno avviate le operazioni per la bancarotta negli Stati Uniti. Nonostante questo e la "precaria condizione finanziaria", la società si impegna a far fronte a tutti gli obblighi nei confronti dei suoi dipendenti, inclusa la distribuzione delle indennità di fine rapporto.

¹⁴ (Cambridge Analytica, 2018): "Negli ultimi mesi siamo stati oggetto di numerose accuse infondate e, nonostante i nostri sforzi di rettifica, siamo stati denigrati per attività che non solo sono legali ma sono anche ampiamente accettate come componente standard della pubblicità online sia nell'arena politica sia in quella commerciale."

2) Sospensione di 200 App: dopo lo scandalo generato dal caso Cambridge Analytica, Facebook sta cercando di monitorare in maniera più efficace le applicazioni di terze parti presenti all'interno del social media. Lo scopo della verifica attuata è stata quella di capire quali app hanno rubato dati degli utenti e sospenderle. Dopo una prima analisi, Facebook ha deciso di sospenderne $\cong 200$, che hanno registrato e condiviso online le informazioni inserite dagli utenti iscritti al social network. Gli stessi responsabili per la sicurezza e privacy hanno dichiarato che si tratta solo di un primo passo verso la protezione degli utenti e che prossimamente verranno attuati altri controlli.

In seguito allo scandalo, nella home degli utenti sono comparsi alcuni avvisi per una gestione più chiara delle impostazioni di privacy, come in Figura 4.1,

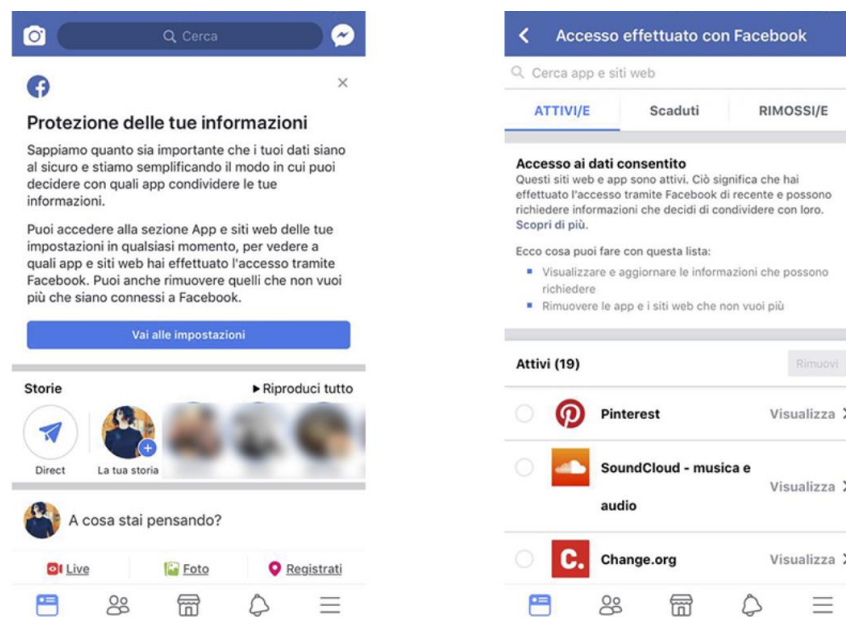


Figura 4.1: Protezione delle informazioni, <https://bit.ly/2tYbU39>

esplicitando il modo per capire quali app e siti siano collegati al proprio profilo e poterli cancellare, oppure come in Figura 4.2, nella quale è presente un avviso che ci permette di verificare se facciamo parte degli utenti profilati da Cambridge Analytica.

Leggi di seguito per sapere se le tue informazioni sono state condivise con Cambridge Analytica tramite l'app "This Is Your Digital Life".



Figura 4.2: Verifica profilazione da parte di Cambridge Analytica, <https://bit.ly/18AOiDE>

- 3) **Clear History, strumento per la cancellazione dei cookies:** durante la conferenza F8 di Facebook, Mark Zuckerberg ha annunciato alcune novità che verranno introdotte sulla piattaforma, tra le quali è presente anche *clear history*, strumento che permette agli utenti di eliminare i dati che Facebook raccoglie da siti e apps, al di fuori del social network. Ogni utente potrà cancellare parte della cronologia di navigazione dall'archivio dati di Facebook. La società è in grado di raccogliere queste informazioni grazie a una serie di plug-in di Facebook che inviano i dati al social network. Principalmente queste informazioni vengono utilizzate per indirizzare annunci pubblicitari mirati agli utenti della piattaforma. La nuova funzione, attualmente in fase di sviluppo, consentirà per la prima volta agli utenti del social di disattivare questa raccolta di dati.

- 4) Il 10 aprile 2018, Facebook ha lanciato il **Data Abuse Bounty**, al fine di premiare gli utenti che segnalano l'uso improprio dei dati da parte di sviluppatori di applicazioni. La ricompensa sarà proporzionale alla mole dell'abuso comunicato.

5) Facebook Container, estensione del browser Firefox



Figura 4.3: Facebook Container Extension, <https://mzl.la/2zhx7KK>

Immediatamente dopo lo scandalo di CA, Mozilla ha rilasciato questa estensione che permette di mantenere l'attività di Facebook separata dal resto della navigazione online, limitando i dati trattenuti dagli altri siti, fin quando non sarà disponibile lo strumento, precedentemente citato, Clear History. Facebook Container rende più difficile al social network raccogliere informazioni sui siti visitati e creare pubblicità ad hoc in base alla propria attività. Come avviene tutto ciò? All'installazione, questa estensione cancella i cookie del social ed effettua il log-out. Al momento dell'accesso, Facebook sarà caricato in una scheda blu, come in Figura 4.4, e tutti i collegamenti verranno aperti in schede separate, esterne al social. Per esempio, i pulsanti *Mi piace* presenti nelle schede al di fuori del contenitore, non funzioneranno e non si potrà effettuare l'autenticazione tramite Facebook, come "Accedi/Registrati con Facebook".

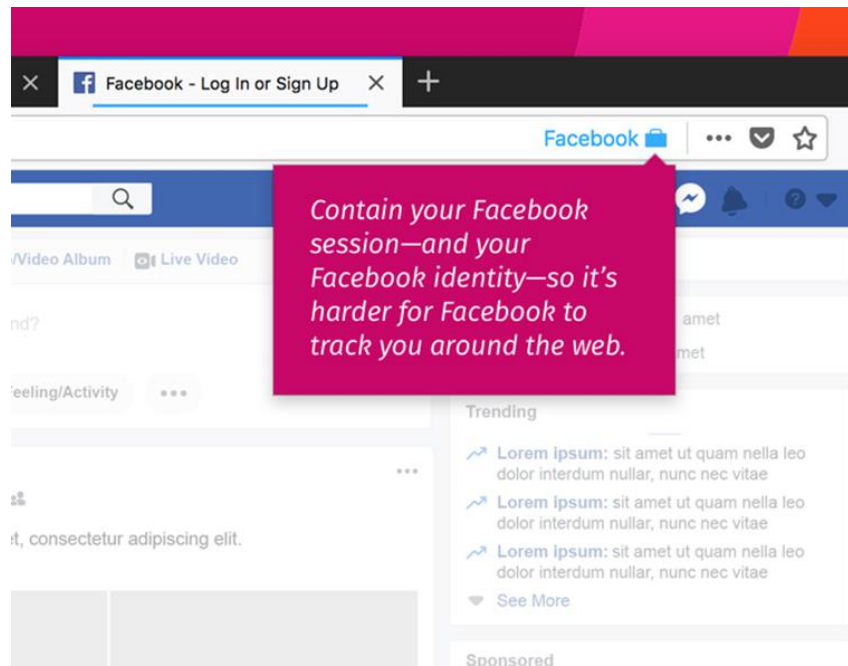


Figura 4.4: Facebook Container, <https://mzl.la/2zhx7KK>

- 6) Google introduce nuove regole per la pubblicità durante le campagne elettorali in USA:** Google, il 5 maggio, ha reso nota la notizia che verrà attuata la verifica degli annunci da parte degli inserzionisti elettorali negli Stati Uniti. In particolare, qualsiasi inserzionista che desideri acquistare spazi pubblicitari a scopo elettorale dovrà dimostrare di essere cittadino o residente legittimo nel Paese, come richiesto dalla legge, fornendo un documento di identità valido. Verrà richiesto, inoltre, che il finanziatore della campagna venga chiaramente indicato.

Google si è assunta pubblicamente l'impegno di rendere trasparente le campagne pubblicitarie legate alle elezioni (anche come risposta al *Russiagate*) che tenterebbero di influenzare le elezioni americane proprio attraverso internet. In arrivo, l'introduzione di un rapporto dettagliato sulla trasparenza degli annunci, che specificherà chi sta acquistando annunci elettorali e quanto sta spendendo. Tutto questo sarà reso disponibile attraverso una libreria aperta e consultabile online.

4.1.1 General Data Protection Regulation

Il Regolamento Ue 2016/679, noto come GDPR (General Data Protection Regulation) è la più grande revisione delle norme sulla privacy online sin dalla nascita di Internet. Il testo, approvato il 14 aprile 2016 dal Parlamento europeo, pubblicato sulla Gazzetta Ufficiale europea il 4 maggio 2016 ed entrato in vigore il 25 maggio dello stesso anno, inizierà ad avere efficacia ed essere direttamente applicabile in tutti gli Stati membri a partire dal 25 maggio 2018. Esso pone le basi per un controllo più efficace dei dati personali. Il GDPR è destinato a tutte quelle aziende che raccolgono e/o elaborano dati personali di cittadini europei, sia online che offline, all'interno o anche fuori dall'Unione europea.

“La definizione presente nell'articolo 4 stabilisce l'oggetto del regolamento: **«dato personale»**: qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»); si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale.”¹⁵

La definizione di dato personale e sensibile, inoltre, risulterà ampliata e includerà non più solo i classici dati sensibili come indirizzo o numero di telefono, ma anche gli identificativi online come cookie, indirizzi IP, geo-localizzazione ed email.

Un deciso cambiamento si ha nei confronti dell'autorizzazione che il soggetto, a cui fanno capo i dati personali, deve dare per un determinato trattamento. L'attuale Codice della Privacy permette, in taluni situazioni, formule per cui la mancanza di opposizione ad un trattamento diventa implicitamente un assenso.

“Il Regolamento Europeo si esprime definendo **«consenso dell'interessato»**: qualsiasi manifestazione di volontà libera, specifica, informata e inequivocabile dell'interessato, con la quale lo stesso manifesta il proprio assenso, mediante dichiarazione o azione positiva inequivocabile, che i dati personali che lo riguardano, siano oggetto di trattamento. Quindi, in ogni caso, l'interessato dovrà esprimersi, in merito al trattamento, fornendo un consenso

¹⁵ (Homerger, 2018)

effettivo ed inequivocabile”¹⁶, ad esempio, con dichiarazione scritta, orale o attraverso mezzi digitali.

Ora ci troviamo così, nella situazione in cui “ogni operazione di comunicazione dovrà quindi far capo ad un preciso consenso formalizzato per ogni controparte presentabile, nel caso l’interessato ne facesse richiesta.”¹⁷ L’interessato ha il diritto di accedere ai dati personali raccolti che lo riguardano e di revocare qualsiasi consenso abbia dato.

Come detta l’articolo 4 del Regolamento, si dovranno definire formalmente due figure *responsabili della privacy*: “il «**titolare del trattamento**»: persona fisica o giuridica, l’autorità pubblica, il servizio o altro organismo che, singolarmente o insieme ad altri, determina le finalità e i mezzi del trattamento di dati personali; quando le finalità e i mezzi di tale trattamento sono determinati dal diritto dell’Unione o degli Stati membri, il titolare del trattamento o i criteri specifici applicabili alla sua designazione possono essere stabiliti dal diritto dell’Unione o degli Stati membri;”¹⁸ il titolare, con un contratto, designa il “«**responsabile del trattamento**»: persona fisica o giuridica, l’autorità pubblica, il servizio o altro organismo che tratta dati personali per conto del titolare del trattamento.”¹⁹

L’idea innovativa introdotta dal GDPR è quella di introdurre un meccanismo di responsabilizzazione dei titolari e dei responsabili del trattamento, per far sì che il Regolamento sia applicato correttamente.

Per far sì che i dati siano trattati in modo lecito e trasparente, la normativa richiede che i titolari e i responsabili debbano, *in primo luogo*, tenere il registro dei trattamenti. I registri da mantenere e conservare sono due: **il registro del titolare del trattamento**, che contiene i dati anagrafici del titolare, del rappresentante e del responsabile alla protezione dati, lo scopo del trattamento, la categoria dell’interessato, a cui fa capo il dato, e le condizioni per la cancellazione automatica del dato in questione; **il registro del responsabile del trattamento**, nel quale sono presenti i dati anagrafici del responsabile del trattamento, la descrizione del

¹⁶, ¹⁷, ¹⁸, ¹⁹ (Homerger, 2018)

trattamento effettuato, facendo riferimento alla categoria dell'interessato ed eventualmente l'enunciazione delle misure di sicurezza avviate. Possono essere conservati in forma cartacea ma anche in forma elettronica, lasciando sempre a disposizione il dato per ispezioni da parte dell'autorità garante.

In secondo luogo, adottare misure di sicurezza, come notificare le violazioni di dati personali all'autorità di controllo, entro 72 ore dalla scoperta, a meno che la violazione non presenti alcun rischio.

“La notifica dovrebbe includere:

- Una descrizione della natura della violazione dei dati personali compresi, ove possibile, le categorie e il numero approssimativo di interessati in questione, nonché le categorie e il numero approssimativo di registrazioni dei dati personali;
- La comunicazione con il nome e i dati di contatto del responsabile della protezione dei dati;
- Una descrizione delle probabili conseguenze della violazione dei dati personali;
- Una spiegazione delle misure adottate (o in via di adozione) da parte del titolare del trattamento per porre rimedio alla violazione o per attenuarne i possibili effetti negativi.”²⁰

È importante sottolineare che, con il nuovo GDPR, i titolari non devono solo garantire l'osservanza delle norme, ma anche dimostrare di metterle in pratica. Per questo motivo si introduce il concetto di «**accountability**», una vera e propria responsabilità sull'attuazione delle misure, che prevede una **valutazione d'impatto della protezione dei dati** (una valutazione del rischio a tutti gli effetti). Correlato ad essa è il «**responsabile della protezione dei dati**» (DPO - Data Protection Officer), una nuova figura che ha il compito di accertarsi che il titolare o il responsabile al trattamento dei dati osservino gli obblighi previsti.”²¹

²⁰, ²¹ (Catania, 2018)

“Il nuovo testo introduce il cosiddetto «**diritto all’oblio**», che consiste, nella facoltà da parte dell’interessato, di ottenere dal titolare del trattamento la cancellazione delle informazioni personali se:

- I dati personali non sono più necessari rispetto alle finalità per le quali sono stati raccolti o altrimenti trattati;
- L’interessato revoca il consenso su cui si basa il trattamento e se non sussiste altro fondamento giuridico per il trattamento;
- L’interessato si oppone al trattamento e non sussiste alcun motivo legittimo prevalente per procedere al trattamento;
- I dati personali sono stati trattati illecitamente;
- I dati personali devono essere cancellati per adempiere un obbligo legale previsto dal diritto dell’Unione o dello Stato membro cui è soggetto il titolare del trattamento;
- I dati personali sono stati raccolti quando l’interessato era minore e dunque non pienamente consapevole dei rischi derivanti dal trattamento.”²²

Tutti i dati in possesso da parte dei social network dovranno essere facilmente accessibili, scaricabili, modificabili, cancellabili o trasferibili a un altro servizio analogo (*portabilità dei dati*);

Cambiano anche le modalità di registrazione e fruizione di molti servizi internet. Così varia anche la visione relativa alla «**profilazione dell’utente**» definita dal Regolamento:

“Qualsiasi forma di trattamento automatizzato di dati personali consistente nell’utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l’affidabilità, il comportamento, l’ubicazione o gli spostamenti di detta persona fisica. Nel caso di marketing diretto, l’interessato avrà sempre diritto di opporsi alle attività di profilazione.”²³

“Il trasferimento di dati personali verso un **Paese terzo** o un’organizzazione internazionale è ammesso solo nel caso in cui sussistano **adeguate condizioni di protezione**. Tale conformità viene valutata dalla Commissione Europea sulla base di alcuni elementi fondamentali - come stato di diritto, rispetto dei diritti umani e delle libertà fondamentali,

²² (Catania, 2018)

²³ (Homerger, 2018)

legislazione generale e settoriale - nonché sugli impegni internazionali assunti dal paese terzo o dall'organizzazione internazionale e sull'esistenza e l'effettivo funzionamento di una o più autorità di controllo indipendenti.”²⁴

Resta, comunque, vietato il trasferimento di dati personali verso Paesi situati al di fuori dell'Unione europea o organizzazioni internazionali che non rispondono agli standard di adeguatezza.

“Un caso a sé riguarda i **dati provenienti da minori**, soggetti di età inferiore a 16 anni (o, in casi specifici, di 13 anni) che *possono essere meno consapevoli dei rischi, delle conseguenze e delle misure di salvaguardia interessate nonché dei loro diritti in relazione al trattamento dei dati personali* e per i quali si parla di **specifica protezione**. Questa condizione speciale - che sul piano dell'autorizzazione si deve tradurre in un intervento del genitore (a meno che non si tratti di servizi di prevenzione o di consulenza forniti direttamente al minore) - dovrebbe riguardare, in particolare, l'utilizzo dei dati personali a fini di marketing o di creazione di profili di personalità o di utente.”²⁵

Sanzioni Privacy 2018 per le violazioni al regolamento GDPR:

- Sanzioni fino a 10 milioni di euro o 2% del fatturato totale annuo di un'impresa per: violazione delle condizioni applicabili al consenso dei minori, trattamento illecito di dati personali che non richiede l'identificazione dell'interessato, mancata o errata comunicazione di un data breach all'autorità nazionale competente, violazione dell'obbligo di nomina del DPO (Data Protection Officer), mancata applicazione di misure di sicurezza;
- Sanzioni fino a 20 milioni di euro o 4% del fatturato totale annuo di un'impresa per: inosservanza di un ordine imposto dall'autorità nazionale competente e trasferimento illecito di dati personali ad un destinatario in un Paese terzo.

In Italia, sarebbe stato auspicabile che il Governo fosse riuscito ad esercitare la delega ad esso assegnata con Legge n. 163/2017 art. 13. Ciò non è avvenuto ed il

²⁴, ²⁵ (Catania, 2018)

Governo entro sei mesi dalla pubblicazione della stessa Legge di delega, avrebbe dovuto adottare un decreto legislativo di adeguamento della normativa italiana al GDPR, con riguardo unicamente alle materie in cui lo stesso GDPR contempla la competenza delle normative nazionali. Purtroppo, come risaputo, la suddetta delega non è stata esercitata nei termini previsti ed è stata concessa una proroga di 3 mesi, che scadrà il 22 agosto 2018. La soluzione, presumibilmente più coerente con il sistema, è che il nostro Codice privacy, almeno nelle parti in contrasto con il GDPR, non possa più essere applicato dopo il 25 maggio.

CONCLUSIONI

Cambridge Analytica era un'azienda che attraverso la scienza del comportamento e l'analisi dei dati ha cercato di sintetizzare la psicologia e l'analisi predittiva basata sui big data, per comprendere il pubblico, in ambito politico, in maniera tale da migliorarne gli effetti della comunicazione. Tutto questo si può sintetizzare in una definizione: *microtargeting comportamentale* (pubblicità altamente personalizzata su ogni singola persona). I responsabili di CA sostenevano di riuscire a far leva non solo sui gusti, ma sulle emozioni degli utenti. Questo approccio alla propaganda elettorale – e in effetti al marketing – sarà l'eredità definitiva di Cambridge Analytica. Luca Passani, chief technology officer presso ScientiaMobil, ha dichiarato:

“Cambridge Analytica ha costruito la prima macchina per manipolare il consenso. Da un certo punto di vista, si possono truccare le elezioni operando i brogli direttamente nella mente degli elettori. Chi controlla la macchina ha a disposizione una dashboard con cui influenzare pesantemente il processo democratico. [...] Questa è una minaccia che i governi di tutti i paesi devono riconoscere e che andrà neutralizzata con leggi e regolamentazioni efficaci.”²⁶

Facebook, Google e altri stanno continuamente spingendo i confini della pubblicità digitale, implementando Intelligenza Artificiale, realtà virtuale, neuromarketing e altre tecniche. Stanno gettando le basi per il mondo di "Internet of Things", dove saremo ulteriormente monitorati e mirati, ovunque andremo e qualunque cosa vorremo fare. Grazie alla raccolta di dati e al costante monitoraggio delle nostre attività, i servizi online che utilizziamo ci offrono *user experience* sempre migliori. Quante volte ci siamo lamentati della scarsa user interface di un sito o un'applicazione, o abbiamo preferito un servizio piuttosto che un altro per la facilità dell'utilizzo dei canali? Ogni sistema per funzionare meglio e per rendere l'esperienza migliore ha studiato i propri utenti con molta attenzione, per capirne la psiche tramite sondaggi, studiandone i comportamenti o attuando ricerche di mercato. Oggi avviene tutto in modo più veloce, più rapido e personale, le compagnie che operano nell' *internet of things* possiedono tutte le nostre informazioni e siamo stati proprio noi a fornirglielo.

²⁶ (Passani, 2018)

Il potere e la precisione dei *big data* tende ad essere spesso sottovalutata: sicuramente, è difficile dire con certezza quale impatto Cambridge Analytica abbia avuto sulle elezioni americane e quanta importanza dare al processo di targeting applicato agli utenti Facebook, tuttavia l'inaspettata vittoria di Trump e il fallimento dei sondaggi atti a prevedere il vincitore, mi fanno sostenere questa tesi.

Inoltre, sebbene non sia mai scoppiato uno scandalo di tale portata, è esattamente questo lo scopo e il tipo di utilizzo per cui Facebook è stato progettato. Tale piattaforma continua ad essere un'azienda di successo finanziario perché, da sempre, ha consentito a terzi di applicare il microtargeting comportamentale ai soggetti presenti su di essa.

A conclusione di questa tesi pongo un quesito finale: in questa era dei big data, il processo democratico e le nostre preferenze, sia politiche che socio-comportamentali, quanto potranno realmente essere manipolate?

BIBLIOGRAFIA

- Sisti A (2017), Digital transformation war. Milano: Franco Angeli;

SITOGRAFIA

- Amato F (2018), Cambridge Analytica, ecco come ha sfruttato i big data per spingere l'elezione di Trump. L'inchiesta di FqMillenium, *Il Fatto Quotidiano*, 18 marzo;
- Apply Magic Sauce, Technical Documentation. Disponibile a: <https://applymagicsauce.com/>
- Cadwalladr C, Graham-Harrison E (2018), Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach, Whistleblower describes how firm linked to former Trump adviser Steve Bannon compiled user data to target American voters, *The Guardian*, 17 marzo;
- Cambridge Analytica (2018), Cambridge Analytica and Scl Elections Commence Insolvency Proceedings and Release Results of Independent Investigation into Recent Allegations. Disponibile a: <https://ca-commercial.com/news/cambridge-analytica-and-scl-elections-commence-insolvency-proceedings-and-release-results-3>;
- Catania R (2018), GDPR: tutto quello che c'è da sapere in 12 punti, L'informativa, il consenso, gli obblighi (e le sanzioni) per i titolari: cosa dice il nuovo regolamento europeo per il trattamento dei dati personali, *Panorama*, 25 maggio;
- Chang A (2018), The Facebook and Cambridge Analytica scandal, explained with a simple diagram, A visual of how it all fits together. They're now shutting down, *Vox*, 2 maggio;
- Crea A (2018), Facebook Clear History, proteggi la privacy in un click, *tom's HARDWARE*, 2 maggio;
- Dopo Cambridge Analytica Facebook sospende 200 app per uso dati, Migliaia quelle indagate, se hanno compiuto abusi saranno bandite, *ANSA*, 14 maggio 2018;

- EUGDPR.org. (2018) GDPR Key Changes. Disponibile a: <https://www.eugdpr.org/>;
- European Commission. (2018) 2018 Reform of EU Data Protection Rules. Disponibile a: https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en;
- Facebook Business. (2014) Learn More About the People that Matter to Your Business with Facebook Audience Insights. Disponibile a: <https://www.facebook.com/business/news/audience-insights>;
- Facebook, scandalo Cambridge Analytica: le cose da sapere, Zuckerberg fa mea culpa. L'Antitrust americana apre un'inchiesta sull'uso dei dati personali da parte del social network. La società di data mining sospende l'amministratore delegato. E in California parte la prima class action. Lo scandalo dei profili rubati a punti, *Lettera43*, 20 marzo 2018;
- Gaggi M, Facebook e Google: la battaglia contro le notizie false in rete, Tagliati i proventi pubblicitari ottenuti dai falsi siti di notizie: le due compagnie della Silicon Valley hanno adottato alcune misure per scoraggiare la diffusione di informazioni non veritiere in rete, *Corriere della Sera*;
- Gazzetta Ufficiale dell'Unione Europea, REGOLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 27 aprile 2016. Disponibile a: <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32016R0679&from=IT>
- Hindman M (2018), How Cambridge Analytica's Facebook targeting model really worked, The researcher who built the model claims it wasn't very accurate on an individual level, *ArsTechnica*, 4 aprile;
- Homberger L (2018), IL GDPR IN 10 PASSI: IL NUOVO REGOLAMENTO EUROPEO SULLA PRIVACY, *CWS Digital solution*. Disponibile a: https://www.cws.it/pdf/assets/0L1nB_II%20GDPR%20in%2010%20passi.pdf
- Killian G (2018), Psychographics: the behavioural analysis that helped Cambridge Analytica know voters' minds, *The Conversation*, 21 marzo;
- Kosinski M, Stillwell D, Graepel T: Private traits and attributes are predictable from digital records of human behavior, *Pnas*, 110 (15) 5802-5805;

- Matz S C, Kosinski M., Nave G, Stillwell D, Psychological targeting as an effective approach to digital mass persuasion, *Pnas*, 114 (48) 12714-12719;
- Mozilla Firefox. (2018) Mantieni la tua privacy, traccia i limiti di Facebook. Disponibile a: <https://www.mozilla.org/it/firefox/facebookcontainer/>;
- Newsroom.fb, Data Abuse Bounty: Facebook Now Rewards for Reports of Data Abuse. Disponibile a: <https://newsroom.fb.com/news/2018/04/data-abuse-bounty/>.
- Venturini M (2018), Cambridge Analytica, vestito blu e messaggio agli utenti: la tattica di Zuckerberg per difendere Facebook, *Il Fatto Quotidiano*, 12 aprile;
- Wakefield J (2018): Cambridge Analytica: Can targeted online ads really change a voter's behaviour?, *BBC*, 30 marzo;
- Wikipedia. Cross-validation (statistics). Disponibile a: [https://en.wikipedia.org/wiki/Cross-validation_\(statistics\)#k-fold_cross-validation](https://en.wikipedia.org/wiki/Cross-validation_(statistics)#k-fold_cross-validation)

FIGURE

- 1.1 http://www.tgcom24.mediaset.it/mondo/cambridge-analytica-la-talpa-dietro-c-era-bannon-class-action-negli-usa_3129619-201802a.shtml
- 1.2 <https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram> (trad. dell'autore)
- 2.1 <http://slideplayer.it/slide/5400561/>
- 2.2 <https://www.enkimd.com/big-five-personality-traits.html>
- 2.3 <http://www.pnas.org/content/110/15/5802> (trad. dell'autore)
- 2.4 <http://www.pnas.org/content/110/15/5802> (trad. dell'autore)
- 2.5 <http://www.pnas.org/content/110/15/5802> (trad. dell'autore)
- 2.6 <http://www.pnas.org/content/110/15/5802> (trad. dell'autore)
- 2.7 <https://applymagicsauce.com/>
- 2.8 <https://applymagicsauce.com/> (trad. dell'autore)
- 2.9 <https://applymagicsauce.com/> (trad. dell'autore)
- 2.10 <https://applymagicsauce.com/> (trad. dell'autore)
- 2.11 <https://applymagicsauce.com/> (trad. dell'autore)
- 2.12 <https://github.com/campsyh/example-prediction-api/blob/master/LICENSE>
- 3.1 Screenshot: <https://www.youtube.com/watch?v=n8Dd5aVXLCc>
- 3.2 <http://www.pnas.org/content/114/48/12714> (trad. dell'autore)
- 3.3 <http://www.pnas.org/content/114/48/12714> (trad. dell'autore)
- 3.4 <http://www.pnas.org/content/114/48/12714> (trad. dell'autore)
- 3.5 <http://www.pnas.org/content/114/48/12714> (trad. dell'autore)
- 3.6 <https://www.facebook.com/business/news/audience-insights>
- 3.7 <https://www.facebook.com/business/news/audience-insights>
- 3.8 <https://www.facebook.com/business/products/ads/ad-targeting>
- 4.1 <https://www.wired.it/internet/social-network/2018/04/18/cambridge-analytica-cancellare-dati-app-facebook-e-piu-facile/>
- 4.2 <https://www.facebook.com/>
- 4.3 <https://www.mozilla.org/it/firefox/facebookcontainer/>
- 4.4 <https://www.mozilla.org/it/firefox/facebookcontainer/>

RINGRAZIAMENTI

Innanzitutto, desidero ringraziare il mio relatore, il Prof. Edoardo Mollona per aver accettato la mia proposta ed aver permesso di svolgere questo lavoro di tesi.

Desidero ringraziare tutti i miei amici e tutte le mie amiche, che sono sempre stati al mio fianco, pronti a festeggiare per ogni esame passato e non. In particolare, Chiara, Elena e Maria Elena.

Desidero ringraziare tutte le persone che sono passate per Righi 32, con le quali ho vissuto 3 anni meravigliosi e ho condiviso le esperienze più belle della mia vita.

Ringrazio Cecilia, Luca e Gianluca, per la pazzia e la sintonia creatasi in casa.

In particolare, un ringraziamento speciale va a Valentina, per la convivenza, l'amicizia colma di amore e complicità e per avermi insegnato che la distanza non compromette i legami più profondi.

Voglio ringraziare Roberta e Massimo, per aver creduto in me giorno dopo giorno e per aver dimostrato un amore senza pari, puro e genuino.

Il ringraziamento più grande va a tutta la mia famiglia, a mio fratello Matteo per essere stato, da sempre, un esempio di vita e saggezza e per avermi fatto capire che gli ostacoli e le sfide più dure si superano in ogni caso. Alla donna più importante della mia vita, mia mamma Anna, una costante, senza la quale non ce l'avrei mai fatta. All'uomo più importante della mia vita, mio papà Luca, alla forza e all'ottimismo che mi ha sempre trasmesso, grazie a te sono la persona che sono ora.

Ringrazio infine Alessandro, per questo amore speciale, inaspettato e colmo di complicità, devo tutto questo anche a te.

Grazie di cuore a tutti.